JMIR Perioperative Medicine

Technologies for pre- and post-operative education, preventative interventions, and clinical care for surgery and anaesthesiology patients, as well as informatics applications in anesthesia, surgery, critical care, and pain medicine

Volume 6 (2023) ISSN: 2561-9128 Editor in Chief: John F Pearson, MD

Contents

Original Papers

A Neural Network Model Using Pain Score Patterns to Predict the Need for Outpatient Opioid Refills Following Ambulatory Surgery: Algorithm Development and Validation (e40455)

Rodney Gabriel, Sierra Simpson, William Zhong, Brittany Burton, Soraya Mehdipour, Engy Said. 4

The Association Between Borderline Dysnatremia and Perioperative Morbidity and Mortality: Retrospective Cohort Study of the American College of Surgeons National Surgical Quality Improvement Program Database (e38462)

Jacob Cole, Krista Highland, Scott Hughey, Brendan O'Shea, Thomas Hauert, Ashton Goldman, George Balazs, Gregory Booth. 17

Toward Enhanced Clinical Decision Support for Patients Undergoing a Hip or Knee Replacement: Focus Group and Interview Study With Surgeons (e36172)

Sabrina Grant, Emma Tonkin, Ian Craddock, Ashley Blom, Michael Holmes, Andrew Judge, Alessandro Masullo, Miquel Perello Nieto, Hao Song, Michael Whitehouse, Peter Flach, Rachael Gooberman-Hill. 27

The Accuracy of Wrist-Worn Photoplethysmogram–Measured Heart and Respiratory Rates in Abdominal Surgery Patients: Observational Prospective Clinical Validation Study (e40474)

Jonna van der Stam, Eveline Mestrom, Jai Scheerhoorn, Fleur Jacobs, Simon Nienhuijs, Arjen-Kars Boer, Natal van Riel, Helma de Morree, Alberto Bonomi, Volkher Scharnhorst, R Bouwman. 63

Prediction of Pelvic Organ Prolapse Postsurgical Outcome Using Biomaterial-Induced Blood Cytokine Levels: Machine Learning Approach (e40402)

Mihyun Waugh, Nicholas Boltin, Lauren Wolf, Jane Goodwin, Patti Parker, Ronnie Horner, Matthew Hermes, Thomas Wheeler, Richard Goodwin, Melissa Moss. 73

Early Warning Scores to Support Continuous Wireless Vital Sign Monitoring for Complication Prediction in Patients on Surgical Wards: Retrospective Observational Study (e44483)

Mathilde van Rossum, Robin Bekhuis, Ying Wang, Johannes Hegeman, Ellis Folbert, Miriam Vollenbroek-Hutten, Cornelis Kalkman, Ewout Kouwenhoven, Hermie Hermens. 83

Efficacy of Electronic Reminders in Increasing the Enhanced Recovery After Surgery Protocol Use During Major Breast Surgery: Prospective Cohort Study (e44139)

Sumeet Gopwani, Ehab Bahrun, Tanvee Singh, Daniel Popovsky, Joseph Cramer, Xue Geng. 103

A Real-Time Mobile Intervention to Reduce Sedentary Behavior Before and After Cancer Surgery: Pilot Randomized Controlled Trial (e41425)

Carissa Low, Michaela Danko, Krina Durica, Julio Vega, Meng Li, Abhineeth Kunta, Raghu Mulukutla, Yiyi Ren, Susan Sereika, David Bartlett, Dana Bovbjerg, Anind Dey, John Jakicic. 111

XSL•FO
RenderX

A Mobile App for Postoperative Pain Management Among Older Veterans Undergoing Total Knee Arthroplasty: Mixed Methods Feasibility and Acceptability Pilot Study (e50116)
Jessica Morgan, Caitlin Rawlins, Steven Walther, Andrew Harvey, Annmarie O'Donnell, Marla Greene, Troy Schmidt. 121

Description of the Content and Quality of Publicly Available Information on the Internet About Inhaled Volatile Anesthesia and Total Intravenous Anesthesia: Descriptive Study (e47714)
Xinwen Hu, Bethany Pennington, Michael Avidan, Sachin Kheterpal, Nastassjia deBourbon, Mary Politi. 128

Perioperative Risk Assessment of Patients Using the MyRISK Digital Score Completed Before the Preanesthetic Consultation: Prospective Observational Study (e39044)
Fabrice Ferré, Rodolphe Laurent, Philippine Furelau, Emmanuel Doumard, Anne Ferrier, Laetitia Bosch, Cyndie Ba, Rémi Menut, Matt Kurrek, Thomas Geeraerts, Antoine Piau, Vincent Minville. 144

Remote Home Monitoring of Continuous Vital Sign Measurements by Wearables in Patients Discharged After Colorectal Surgery: Observational Feasibility Study (e45113)
Jobbe Leenen, Vera Ardesch, Gijsbert Patijn. 160

Reducing Alcohol Use Before and After Surgery: Qualitative Study of Two Treatment Approaches (e42532)
Lyndsay Chapman, Tom Ren, Jake Solka, Angela Bazzi, Brian Borsari, Michael Mello, Anne Fernandez. 172

Dashboard of Short-Term Postoperative Patient Outcomes for Anesthesiologists: Development and Preliminary Evaluation (e47398)
Rama Sreepada, Ai Chang, Nicholas West, Jonath Sujan, Brendan Lai, Andrew Poznikoff, Rebecca Munk, Norbert Froese, James Chen, Matthias Görges. 182

An Ensemble Learning Approach to Improving Prediction of Case Duration for Spine Surgery: Algorithm Development and Validation (e39650)
Rodney Gabriel, Bhavya Harjai, Sierra Simpson, Austin Du, Jeffrey Tully, Olivier George, Ruth Waterman. 195

Assessing Barriers to Implementation of Machine Learning and Artificial Intelligence–Based Tools in Critical Care: Web-Based Survey Study (e41056)
Eric Mlodzinski, Gabriel Wardi, Clare Viglione, Shamim Nemati, Laura Crotty Alexander, Atul Malhotra. 208

Temporal Generalizability of Machine Learning Models for Predicting Postoperative Delirium Using Electronic Health Record Data: Model Development and Validation Study (e50895)
Koutarou Matsumoto, Yasunobu Nohara, Mikako Sakaguchi, Yohei Takayama, Syota Fukushige, Hidehisa Soejima, Naoki Nakashima, Masahiro Kamouchi. 221

A New Index for the Quantitative Evaluation of Surgical Invasiveness Based on Perioperative Patients' Behavior Patterns: Machine Learning Approach Using Triaxial Acceleration (e50188)
Kozo Nakanishi, Hidenori Goto. 235

Teaching Basic Surgical Skills Using a More Frugal, Near-Peer, and Environmentally Sustainable Way: Mixed Methods Study (e50212)
Ben Smith, Christopher Paton, Prashanth Ramaraj. 255

Survey of the Impact of Decision Support in Preoperative Management of Anemia (i-Anemia): Survey Study (e49186)
Gaëtan Mignanelli, Richard Boyer, Nicolas Bonifas, Emmanuel Rineau, Yassine Moussali, Morgan Le Guen. 265

XSL•FO
RenderX

Viewpoint

Patient Safety of Perioperative Medication Through the Lens of Digital Health and Artificial Intelligence (e34453)
Jiancheng Ye. 47

Review

The Role of Wearable Technology in Measuring and Supporting Patient Outcomes Following Total Joint Replacement: Review of the Literature (e39396)
Gregory Iovanel, David Ayers, Hua Zheng. 54

Editorial

JMIR Perioperative Medicine: A Global Journal for Publishing Interdisciplinary Innovations, Research, and Perspectives (e54344)
Nidhi Rohatgi. 252

XSL•FO
RenderX

Original Paper

A Neural Network Model Using Pain Score Patterns to Predict the Need for Outpatient Opioid Refills Following Ambulatory Surgery: Algorithm Development and Validation

Rodney Allanigue Gabriel[1], MD, MAS; Sierra Simpson[1], PhD; William Zhong[1]; Brittany Nicole Burton[2], MD; Soraya Mehdipour[1], MD; Engy Tadros Said[1], MD

[1]Department of Anesthesiology, University of California San Diego, La Jolla, CA, United States

[2]Department of Anesthesiology and Perioperative Medicine, University of California Los Angeles, Los Angeles, CA, United States

Corresponding Author:
Rodney Allanigue Gabriel, MD, MAS
Department of Anesthesiology
University of California San Diego
9400 Campus Pt Dr
La Jolla, CA, 92037
United States
Phone: 1 8586637747
Email: ragabriel@health.ucsd.edu

Abstract

Background: Expansion of clinical guidance tools is crucial to identify patients at risk of requiring an opioid refill after outpatient surgery.

Objective: The objective of this study was to develop machine learning algorithms incorporating pain and opioid features to predict the need for outpatient opioid refills following ambulatory surgery.

Methods: Neural networks, regression, random forest, and a support vector machine were used to evaluate the data set. For each model, oversampling and undersampling techniques were implemented to balance the data set. Hyperparameter tuning based on k-fold cross-validation was performed, and feature importance was ranked based on a Shapley Additive Explanations (SHAP) explainer model. To assess performance, we calculated the average area under the receiver operating characteristics curve (AUC), F1-score, sensitivity, and specificity for each model.

Results: There were 1333 patients, of whom 144 (10.8%) refilled their opioid prescription within 2 weeks after outpatient surgery. The average AUC calculated from k-fold cross-validation was 0.71 for the neural network model. When the model was validated on the test set, the AUC was 0.75. The features with the highest impact on model output were performance of a regional nerve block, postanesthesia care unit maximum pain score, postanesthesia care unit median pain score, active smoking history, and total perioperative opioid consumption.

Conclusions: Applying machine learning algorithms allows providers to better predict outcomes that require specialized health care resources such as transitional pain clinics. This model can aid as a clinical decision support for early identification of at-risk patients who may benefit from transitional pain clinic care perioperatively in ambulatory surgery.

(JMIR Perioper Med 2023;6:e40455) doi:10.2196/40455

KEYWORDS

opioids; ambulatory surgery; machine learning; surgery; outpatient; pain medication; pain; pain management; patient needs; predict; algorithms; clinical decision support; pain care

Introduction

Opioids play an essential role in acute perioperative pain management. Increased attention to pain management as a quality metric has brought to light an overuse of prescription opioids contributing to an epidemic across the United States. The United States has had increased opioid prescriptions filled in the immediate postoperative period; a study reported that the mean dose of opioids prescribed for most surgical procedures in the United States was higher than that prescribed in other

XSL•FO
RenderX

countries [1]. Opioid-related deaths made up one of the largest causes of preventable deaths in the United States, which costed an estimated US $78.5 billion to US $1.02 trillion to the US health care system [2,3].

Persistent opioid prescribing is often postsurgical [4], in which as many as 3% of opioid-naive patients required opioids for more than 90 days after a major elective surgery [5]. One potential service that may help curb outpatient opioid use after surgery is the transitional pain clinic, which consists of a team of providers who implement multidisciplinary opioid-sparing approaches such as pharmacological, nonpharmacological, and psychological interventions with the goal of weaning patients from opioids postoperatively as outpatients [6,7]. Transitional pain clinics have been shown to reduce opioid use postoperatively, symptoms of anxiety and depression, pain catastrophizing, and postsurgical pain [8,9]. Given the increased resources required to provide this type of service, not all surgical patients may realistically receive postoperative care from transitional clinics. Currently, the criteria for recommendation to transitional pain services for surgical patients are not uniformly defined; thus, accurate predictive methods for patients who may benefit from transitional pain clinics are needed. Less work has been done on patients undergoing ambulatory surgery and on the identification of patients who may likely require more opioids as an outpatient. In such populations, machine learning may be used to identify postoperative opioid use in the recovery room [10]. In addition, some studies have described the risk factors for using outpatient opioids after ambulatory surgery [11-13].

The objective of this study was to develop machine learning–based predictive models that may aid in the identification of patients likely to require opioid refills after their initial discharge prescription. Specifically, pain score patterns were incorporated (ie, trends in reported pain scores in the recovery room) into the models. We focused on patients who underwent ambulatory surgery, which included orthopedic surgery (eg, joint arthroscopy, forearm/hand surgery), nonmastectomy breast surgery, urology (eg, cystoscopy), minimally invasive surgery (eg, cholecystectomy, hernia repair), colorectal surgery (eg, hemorrhoidectomy), and gynecology (dilation and curettage/evacuation, hysteroscopy). We hypothesized that the use of neural networks that incorporate various features, including recovery room pain phenotypes, may identify patients at higher risk. The pain phenotypes included patterns in patient-reported pain scores in the recovery room, including trajectory of pain and median and maximum pain scores.

Methods

Study Population

Data were retrospectively collected from the electronic medical records of patients who underwent outpatient surgery from January to July 2020 at a single ambulatory surgery center. The outpatient surgeries included in the analysis included orthopedic surgery (eg, joint arthroscopy, forearm/hand surgery), nonmastectomy breast surgery, urology (eg, cystoscopy), minimally invasive surgery (eg, cholecystectomy, hernia repair), colorectal surgery (eg, hemorrhoidectomy), and gynecology (dilation and curettage/evacuation, hysteroscopy).

Ethics Approval

Our institutional review board (Human Research Protections Program) waived the consent requirement and approved this retrospective study (protocol 210099).

Primary Outcome and Features

The primary outcome of interest was a binary variable (response range, 0 or 1), in which 0 was defined as “no opioid refill” and 1 was defined as the patient “needed to refill their outpatient opioid prescriptions” within 2 weeks after surgery (no opioid refill vs opioid refill). This was captured retrospectively from the electronic medical record by review of the following: (1) any telephone note describing patient calling in for opioid renewal; (2) any progress/office visit note from primary care provider, pain medicine specialist, or surgical provider describing the need for opioid refill; or (3) any renewal order in the medication list for opioids within this time frame. On postanesthesia care unit (PACU) discharge, all patients were prescribed up to 5 days of opioids. For perioperative multimodal analgesia, all patients received preoperative acetaminophen unless contraindicated. For a subset of surgical procedures, regional nerve blocks were routinely offered preoperatively (ie, shoulder, hand/forearm, and knee surgeries). Intraoperatively, patients may have received fentanyl, hydromorphone, ketamine, ketorolac, and dexmedetomidine at the discretion of the anesthesiologist. In the PACU, patients were given oxycodone, fentanyl, and hydromorphone, as needed.

Features that were integrated into the model were collected retrospectively from the electronic medical record system. The data included age (years), sex (male vs female), body mass index (kg/m^2), English-speaking, comorbidities, regional nerve block performance, general anesthesia, intraoperative ketamine, intraoperative total intravenous anesthesia, opioid consumption, and pain scores (11-point numeric rating scale [NRS] from 0 to 10). These features were included, as they were determined to be relevant to postoperative opioid use based on clinical judgement and previous research [14,15]. Opioid consumption, defined as total opioids consumed intraoperatively and in the PACU, was measured in intravenous morphine equivalents (MEQ). Pain scores were captured as preoperative pain score, median pain score in the PACU, maximum pain score in the PACU, and slope of pain score trajectory in the PACU. Preoperative pain scores were collected by nurses upon arrival for preoperative check-in. PACU pain scores were captured every 5-15 minutes and recorded in the electronic medical record. A negative value for the pain score slope was defined as an overall decrease in pain scores throughout the PACU stay. A positive value for the pain score slope was defined as an overall increase in the pain scores throughout the PACU stay. A zero value of the pain score slope was defined as no change in the overall trend of pain scores throughout the PACU stay.

Statistical Analysis

Python (v3.10.1) was used for all statistical analyses. Patient and surgical characteristics were compared with chi-squared test (categorical) and Wilcoxon rank sum test (continuous). A

XSL•FO
RenderX

generalized linear mixed model fit by maximum likelihood was implemented to model the features to the primary outcome of opioid refill. The random effect in this model was the surgical procedure. All features were included in the model, and their association with the outcome was reported by their respective odds ratios (ORs), 95% CI, and *P* values. A neural network model to predict the need for opioid refills following surgery was constructed. Logistic regression, random forest, and support vector machine classifiers were implemented for comparison. For all models, patient data were divided into training and test data sets with a 70:30 split by using a stratified randomized splitter—the train_test_split method from the sci-kit learn library. K-fold cross-validation on the training set was used to tune the hyperparameters and to optimize oversampling techniques as well as to calculate the average sensitivity, specificity, F1-score, and area under the curve (AUC) for the receiver operating characteristic curve. The final version of each model was then validated on the test set and the AUC was reported. Feature importance from the neural network was ranked based on Shapley Additive Explanations (SHAP).

Data Balancing

Synthetic Minority Oversampling Technique (SMOTE) for Nominal and Continuous algorithm and random undersampling were both implemented using the imblearn library [16]. These tools were used to achieve a balanced class distribution with minimal difference between positive and negative outcomes. A data set with a large difference between positive and negative outcomes was considered unbalanced and may make it difficult for predictive machine learning models to draw useful conclusions, given the uneven classification of data.

Random undersampling of the majority outcome is frequently used to reduce the impact of imbalanced data sets; however, SMOTE oversamples were used to create synthetic minority class examples to balance the minority class with the majority class. SMOTE uses samples from the minority class and a set number of nearest neighbors—in this case, 5—to generate synthetic cases from the sample class. Combining the 2 techniques as outlined yielded positive outcomes. Both techniques were only applied on the training set. Different combinations for proportions of minority to majority class were analyzed, ranging from 0.25 to 1.00. After performing k-fold cross-validation, the parameter “sampling_strategy” for the SMOTE class from imblearn was optimal when set to 0.24 and the parameter “sampling_strategy” for the RandomUnderSampler class from imblearn was optimal when set to 0.94. Optimal results were based on which hyperparameters produced the highest performance metrics for the model (eg, AUC, F1-score, sensitivity, specificity).

Machine Learning Models

Four different machine learning classification models were evaluated: neural network, logistic regression, random forest classifier, and support vector classifier. For each model, the following sampling methods were compared: oversampling the training set via SMOTE, undersampling the majority class in the training set, a combination of SMOTE and undersampling of the majority classes, and no oversampling or undersampling technique. The results from the sampling method that provided the optimal results were reported. For each model, all features were included as inputs. One-hot encoding was used for categorical features.

Multilayer Perceptron Neural Network

Using the Keras interface for the TensorFlow library, a shallow feedforward neural network was constructed. The rectified linear unit function was used as the activation function. The final output layer used the sigmoid activation function, and the overall model used the Adam optimizer. Repeated k-fold cross-validation was used to tune the hyperparameters to find the optimal parameter values for the number of hidden layers (1), number of neurons per hidden layer (100), maximum number of iterations (300), batch size (16), and learning rate (0.0001).

Logistic Regression

The logistic regression classifier predicts the probabilities of the different outcome possibilities based on the input. A newton-cg solver regression model was implemented without specifying individual class weights. This model provided a baseline score and helped make the case for improvement over the evaluation metrics. Repeated k-fold cross-validation was used to tune the hyperparameters to find the optimal parameter value for *C* (the strength of the regularization is inversely proportional to *C*), which was 0.3.

Random Forest Classifier

The random forest is an ensemble approach, which has been proven effective for a variety of classification problems. To tune the hyperparameters, we performed repeated k-fold cross-validation to find the optimal parameter values for maximum depth (75), minimal samples required to be at the leaf node (4), minimal samples required to split an internal node (4), and number of estimators (100) (ie, number of trees).

Support Vector Classifier

A support vector classifier maps the data onto an n-dimensional space (n being the number of features) and then identifies the hyperplane decision boundary that best separates the data into 2 classes by maximizing the distance between the hyperplane and the nearest data point in either class. K-fold cross-validation was used to tune the hyperparameters to identify the optimal parameter value for *C*, which was 130.

Performance Metrics

The primary performance metric of interest was the AUC for the receiver operating characteristic curve. In addition, we reported F1-scores, sensitivity, and specificity.

K-Fold Cross-Validation

To effectively tune the hyperparameters of the models, stratified k-fold cross-validation was implemented on the training set to observe the sensitivity, specificity, F1-score, and AUC scores, for 10 splits. For each iteration, the data set was split into 10 groups (folds). One fold served as the test set with the other 9 serving as the training set. When assessing the effectiveness of SMOTE and random undersampling, only the training folds were changed. This process was repeated until each fold served as the test set once. Every model exhibited improved

XSL•FO
RenderX

performance metrics when SMOTE and random undersampling were applied.

Results

Study Cohort Characteristics

There were 1333 patients and 28 unique surgical procedures included in the final analysis, and 144 (10.8%) patients refilled their opioid prescription within 2 weeks after outpatient surgery. Univariate analysis revealed that patients who required opioid refills were more likely to be smokers (32/144, 22.2% vs 156/1189, 13.1%, respectively; P=.005) and had a regional nerve block performed (87/144, 60.4% vs 440/1189, 37%, respectively; P<.001). Those who required opioid refills had higher total perioperative opioid consumption (intraoperative and PACU opioid use; P<.001), preoperative pain scores (P<.001), maximum PACU pain scores (P<.001), and median PACU pain scores (P<.001). Table 1 lists the differences between the opioid refill and non–opioid refill cohorts represented in the study population in order to provide information regarding the baseline characteristics. All surgical procedures included in our analysis are listed in Table 1.

XSL•FO
RenderX

Table 1. Baseline characteristics of the study cohorts.[a]

Feature	No opioid refill (n=1189)	Required opioid refill (n=144)	*P* value
Surgical procedure, n (%)			<.001
Arthrodesis (finger)	10 (0.8)	3 (2.1)	
Arthroscopy (hip)	36 (3)	5 (3.5)	
Arthroscopy (knee)	99 (8.3)	12 (8.3)	
Arthroscopy (shoulder)	61 (5.1)	19 (13.2)	
Arthroscopy (shoulder, with rotator cuff repair)	45 (3.8)	13 (9)	
Arthroscopy (wrist)	9 (0.8)	4 (2.8)	
Breast lumpectomy	56 (4.7)	7 (4.9)	
Transperineal prostate biopsy	17 (1.4)	0 (0)	
Laparoscopic cholecystectomy	29 (2.4)	1 (0.7)	
Cystoscopy	43 (3.6)	0 (0)	
Dilation and curettage of the uterus	76 (6.4)	0 (0)	
Dilation and evacuation of the uterus	36 (3)	0 (0)	
Anorectal examination under anesthesia	29 (2.4)	0 (0)	
Condyloma excision	13 (1.1)	3 (2.1)	
Lesion excision of the head and neck	31 (2.6)	1 (0.7)	
Lesion excision of the upper extremity	78 (6.6)	4 (2.8)	
Extracorporeal shockwave lithotripsy	20 (1.7)	1 (0.7)	
Anal fistulectomy	73 (6.1)	1 (0.7)	
Hemorrhoidectomy	29 (2.4)	8 (5.6)	
Inguinal herniorrhaphy	16 (1.3)	2 (1.4)	
Hysteroscopy	47 (4)	0 (0)	
Incision and drainage of the upper extremity	12 (1)	0 (0)	
ORIF[b], distal radius fracture	71 (6)	15 (10.4)	
ORIF, scaphoid fracture	11 (0.9)	0 (0)	
ORIF, hand	70 (5.9)	10 (6.9)	
Ligament reconstruction with tendon interposition, upper extremity	30 (2.5)	9 (6.3)	
Open carpal tunnel release	50 (4.2)	2 (1.4)	
Arthroscopic anterior crucial ligament repair	92 (7.7)	24 (16.7)	
Patient characteristics			
Age (years), median (quartiles)	43 (32, 60)	46.5 (33.0, 58.25)	.67
Male sex, n (%)	551 (46.3)	70 (48.6)	.59
Body mass index (kg/m^2), median (quartiles)	26.1 (23.0, 30.2)	26.3 (23.9, 31.6)	.14
English speaker, n (%)	1106 (93)	129 (89.6)	.19
Comorbidities, n (%)			
Preoperative opioid use	55 (4.6)	8 (5.6)	.77
Diabetes mellitus	87 (7.3)	13 (9)	.57
Alcohol use	53 (4.5)	7 (4.9)	.99
Active smoker	156 (13.1)	32 (22.2)	<.001
Substance abuse	43 (3.6)	10 (6.9)	.09
Hypertension	255 (21.4)	34 (23.6)	.63

XSL•FO
RenderX

Feature	No opioid refill (n=1189)	Required opioid refill (n=144)	*P* value
Depression/anxiety	246 (20.7)	37 (25.7)	.21
Coronary artery disease	35 (2.9)	3 (2.1)	.78
Dementia	10 (0.8)	1 (0.7)	.99
Renal insufficiency	40 (3.4)	4 (2.8)	.91
Chronic obstructive pulmonary disease	18 (1.5)	2 (1.4)	.99
Asthma	116 (9.8)	16 (11.1)	.71
Obstructive sleep apnea	84 (7.1)	9 (6.3)	.95
Chronic pain	114 (9.6)	18 (12.5)	.34
Congestive heart failure	9 (0.8)	0 (0)	.61
Anesthesia/perioperative medications			
Regional block performed, n (%)	440 (37)	87 (60.4)	<.001
General anesthesia, n (%)	783 (65.9)	107 (74.3)	.05
Intraoperative ketamine used, n (%)	55 (4.6)	9 (6.3)	.51
Total intravenous anesthetic, n (%)	43 (3.6)	6 (4.2)	.92
Total (intraoperative and PACU[c]) opioid consumption (MEQ[d]), median (quartiles)	10 (0, 17)	12.8 (6.9, 26)	<.001
Pain score (numeric rating scale 0-10), median (quartiles)			
Preoperative pain score	0 (0, 3)	0 (0, 6)	<.001
Maximum PACU score	2 (0, 6)	6 (0, 8)	<.001
Median PACU pain score	0 (0, 4)	3 (0, 6)	<.001
Slope of PACU pain	0 (–0.14, 0)	0 (–0.46, 0)	<.001

[a]*P* values were calculated by chi-squared and Wilcoxon rank sum tests for categorical and continuous variables, respectively.

[b]ORIF: open reduction and internal fixation.

[c]PACU: postanesthesia care unit.

[d]MEQ: morphine equivalents.

Mixed Effects Logistic Regression Model

The need for opioid refill within 2 weeks after ambulatory surgery was modeled utilizing a mixed effects logistic regression analysis fit by maximum likelihood, in which the random effect was the surgical procedure (Table 2). Features that were statistically significantly associated with higher odds of need for opioid refill were active smokers (OR 1.99, 95% CI 1.19-3.31; *P*=.009), substance abuse history (OR 2.34, 95% CI 1.02-5.37; *P*=.04), regional block performed (OR 2.81, 95% CI 1.62-4.88; *P*<.001), total opioid consumption (MEQ mg) intraoperatively and in PACU (OR 1.03, 95% CI 1.01-1.06; *P*=.008), and median PACU pain score (NRS) (OR 1.19, 95% CI 1.04-1.36; *P*=.01). A feature that was significantly associated with decreased odds of opioid refill was English-speaking patients (OR 0.47, 95% CI 0.24-0.93; *P*=.03).

XSL•FO
RenderX

Table 2. Mixed effects logistic regression modeling of refills.[a]

Feature	Odds ratio (95% CI)	*P* value
Age (years)	0.99 (0.98-1.01)	.85
Male sex	0.96 (0.63-1.46)	.84
Body mass index (kg/m^2)	1.01 (0.98-1.01)	.48
English speaker	0.47 (0.24-0.93)	.03
Comorbidities		
Preoperative opioid use	0.70 (0.25-1.98)	.51
Diabetes mellitus	1.44 (0.67-3.09)	.36
Alcohol use	0.82 (0.33-2.03)	.66
Active smoker	1.99 (1.19-3.31)	.009
Substance abuse	2.34 (1.02-5.37)	.04
Hypertension	1.26 (0.73-2.18)	.41
Depression/anxiety	1.07 (0.66-1.73)	.79
Coronary artery disease	1.23 (0.30-5.00)	.78
Dementia	1.33 (0.09-18.80)	.84
Renal insufficiency	0.79 (0.24-2.58)	.99
Chronic obstructive pulmonary disease	0.99 (0.19-5.17)	.99
Asthma	0.99 (0.53-1.86)	.99
Obstructive sleep apnea	0.62 (0.26-1.44)	.27
Chronic pain	1.31 (0.62-2.79)	.48
Congestive heart failure	0	.99
Anesthesia/perioperative medications		
Regional block performed	2.81 (1.62-4.88)	<.001
General anesthesia	1.05 (0.56-1.94)	.88
Intraoperative ketamine used	1.17 (0.51-2.65)	.71
Total intravenous anesthetic	1.21 (0.46-3.18)	.71
Total (intraoperative and PACU[b]) opioid consumption (MEQ[c])	1.03 (1.01-1.06)	.008
Pain (numeric rating scale 0-10)		
Preoperative pain score	0.95 (0.86-1.05)	.35
Maximum PACU score	1.02 (0.92-1.13)	.67
Median PACU pain score	1.19 (1.04-1.36)	.01
Slope of PACU pain	0.78 (0.56-1.09)	.16

[a]Results of mixed effects logistic regression modeling need for opioid refill after ambulatory surgery. The random effect in this model was the surgical procedure.

[b]PACU: postanesthesia care unit.

[c]MEQ: morphine equivalents.

Neural Network Approach to Predicting Opioid Refills

Hyperparameter tuning via grid search cross-validation was implemented to identify the best architecture of the multilayer perceptron neural network, which consisted of 1 hidden layer, 100 neurons within the hidden layer, 300 maximum iterations for learning, a batch size of 16, and a learning rate of 0.0001. Based on this architecture, the average AUC calculated from k-fold cross-validation was 0.71 (95% CI 0.68-0.74). The final model was then validated on the test set, which yielded an AUC of 0.75 (Figure 1). The features with the highest impact on model output for the neural network based on the absolute SHA*P* values were performance of a regional nerve block, maximum pain score in the PACU, median pain score in the PACU, active smoking history, and total opioid consumption (intraoperative and PACU) (Figure 2).

XSL•FO
RenderX

Next, various other machine learning–based models were implemented to predict the need for opioid refills after ambulatory surgery. Based on k-fold cross-validation, the average AUCs from models with optimized hyperparameters were identified for support vector machine (0.64, 95% CI 0.57-0.71), random forest (0.66, 95% CI 0.60-0.71), and logistic regression (0.69, 95% CI 0.66-0.74) (Table 3). The final models for each machine learning approach were then validated on a separate test set when SMOTE was not applied versus when SMOTE was applied—the calculated AUCs were identified for support vector machine (0.65), random forest (0.68), and logistic regression (0.73). SMOTE improved the performance of each model.

Figure 1. The calculated area under the curve of each machine learning model when trained on 70% of the data and validated on the remaining 30%. The models predict the need for opioid refill within 2 weeks following ambulatory surgery. AUC: area under the curve.

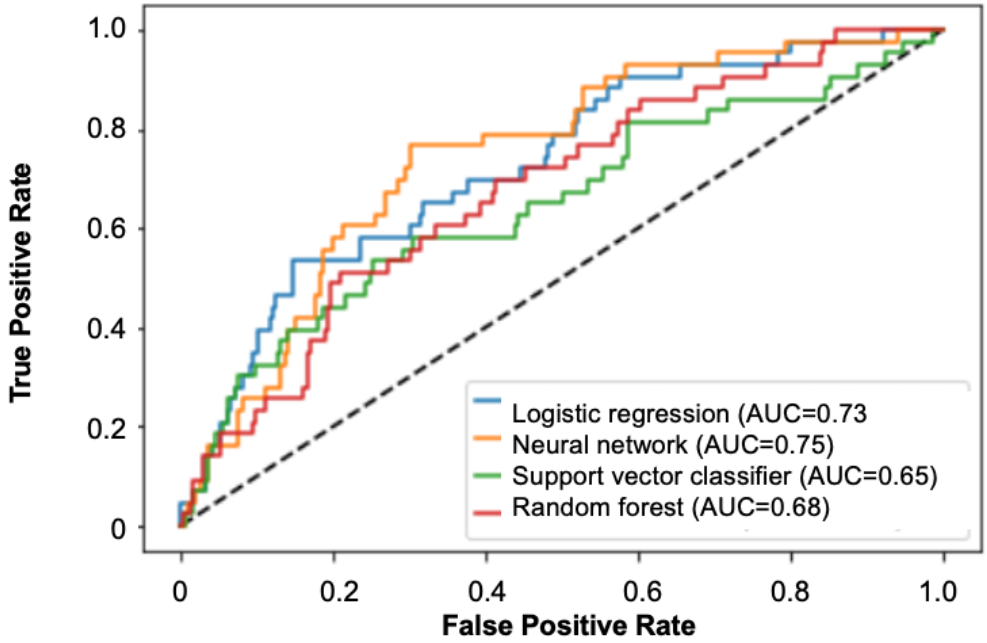
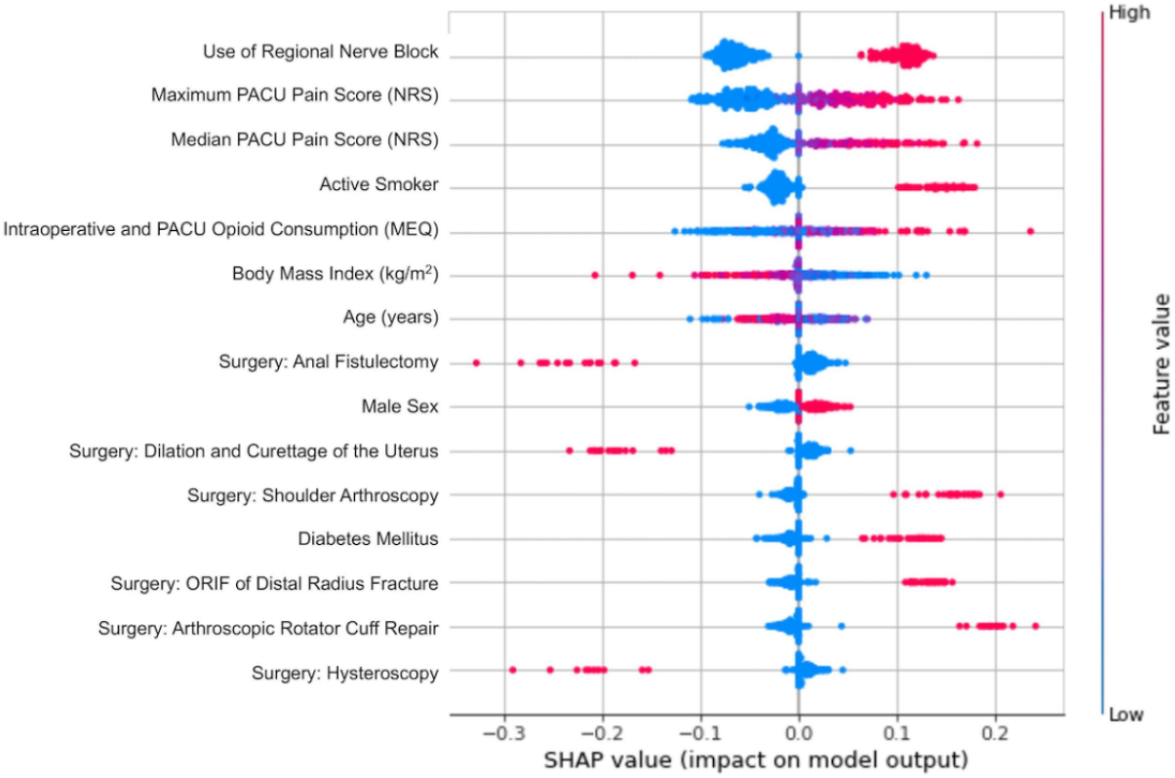


Figure 2. Shapley Additive Explanations (SHAP) values (impact on model output) of the top features used in the neural network predicting the need for opioid refill within 2 weeks following ambulatory surgery. MEQ: morphine equivalents; NRS: numeric rating scale; ORIF: open reduction and internal fixation; PACU: postanesthesia care unit.



XSL•FO
RenderX

Table 3. Model performance for each machine learning model calculated by k-fold cross-validation.

Classification model	Area under the curve		F1-score		Sensitivity		Specificity	
	No SMOTE[a]	SMOTE	No SMOTE	SMOTE	No SMOTE	SMOTE	No SMOTE	SMOTE
Logistic regression	0.516	0.698	0.033	0.335	0.040	0.673	0.992	0.722
Neural network	0.509	0.711[b]	0.067	0.347[b]	0.039	0.693[b]	0.979	0.726
Support vector classifier	0.500	0.643	0.001	0.291	0.002	0.602	0.999[b]	0.684
Random forest	0.500	0.655	0.002	0.317	0.001	0.524	0.999[b]	0.786

[a]SMOTE: Synthetic Minority Oversampling Technique.

[b]Represents the best performance for the given metric.

Discussion

Principal Findings

We demonstrated that a shallow feedforward neural network and other machine learning approaches that integrated pain score patterns had adequate performance to predict the need for opioid refills within 2 weeks following ambulatory surgery. The features with the highest impact on model output were active smoking history, intraoperative opioid consumption, PACU opioid consumption, regional nerve block utilization, as well as maximum and median pain scores in the PACU. The importance of pain score patterns (ie, median and maximum pain scores) in predicting opioid refills is interesting and highlights the association of PACU analgesia and opioid consumption with the requirement for more opioids following the initial prescription. This neural network may be useful in identifying patients at risk who require a longer duration of opioid use so that the limited hospital resources can be better utilized in a precise manner.

Comparison to Prior Work

Previous studies have reported the utilization of machine learning for predicting postoperative opioid use in ambulatory surgery [10,17]. Nair et al [10] reported the accuracies of regression, naïve Bayes, neural networks, random forest, and extreme gradient boosting in predicting postoperative opioid use in the recovery room and showed that random forest performed best when using only preoperative features. Anderson et al [17] utilized models to predict prolonged opioid use specifically in patients who underwent anterior cruciate ligament repair by using regression, Bayesian belief network, gradient boosting, and random forest. They found that gradient boosting was able to achieve an AUC of 0.77. In our study, we reviewed multiple types of ambulatory surgeries and focused on predicting the need for additional outpatient opioid refills weeks after surgery. Four computational approaches were used to determine the best model for our data set, and all had similar performances. Random forest, logistic regression, and support vector machine tools did not perform as well as the neural network, though the random forest model had increased specificity compared to the neural network. Both the support vector classifier and neural network can increase the dimensionality of the data to find a solution, but given the time and training, neural networks usually outperform support vector classifiers. Random forest and neural networks approach data inversely, as random forest decision trees are independent and neural network neurons are dependent on other neurons. Logistic regression is the standard approach but often does not perform well in multidimensional data sets. By surveying multiple models, the benefits of each can be identified and evaluated to improve the validity of the predicted features [18].

Opioids remain the cornerstone for acute postoperative pain management, and the perioperative period is often the patients' first introduction to prescription opioids. Our study's patient cohort was primarily opioid-naïve; only 4.6% (55/1189) of the patients in the non–refill group and 5.6% (8/144) of the patients in the refill group reported preoperative opioid use. Studies have shown surgical procedure as an independent risk factor for prolonged opioid use [4,19,20]. Other risk factors include preoperative opioids, tobacco use, gender, and mood disorders [21-26]. Although efforts are in place to standardize postoperative opioid prescriptions per surgical procedure [27], there continues to be a wide variety in the amount and duration of opioids prescribed and often in excess [1,28-31].

An estimated 67%-92% of the prescribed opioids for postoperative pain remain unused [1,32], leaving great potential for diversion and misuse. An increasing number of heroin users reported first being introduced to opioids via prescription and then resorting to heroin for cost and availability factors [33]. Likewise, Bartels et al [34] report that 80% of the opioid prescriptions remain unused with limited and challenging disposal options. Similarly, a 2017 systematic review reports that patients took only 29%-58% of the prescribed opioid pills [32]. Over the decades, we have learned that excess opioids do not necessarily reduce persistent postsurgical pain or any other pain-related outcomes [35]. As Porter and colleagues [36] demonstrated, sometimes less is more—patient-centered opioid discharge prescription guidelines satisfied 93% of the patients, with 99% in the 0 morphine milligram equivalents group [36]. Although there may be some procedures that do not require postoperative opioids, we must also find a balance and prescribe opioids as necessary to meet individual patients' pain needs [37]. For these reasons, risk stratification can be a helpful tool for guiding the process of postoperative opioid prescribing.

The use of regional anesthesia was associated with opioid refilling. It is important to note that there is no causality that may be drawn from these results but rather an association. It may be that the use of regional anesthesia was associated with surgeries that were more painful in nature, and despite pain

XSL•FO
RenderX

scores being likely lower in the PACU, this group would more likely require additional opioids as outpatients when compared to other surgical procedures that are less likely to receive regional anesthesia for pain management. Other potential limitations include the variability in surgery type, which may range in pain level, both during surgery and during recovery, as well as the subjective nature of pain scores. Despite these limitations, the features that have been identified are actionable and trackable in future studies.

Limitations

There are several limitations in this study—mainly due to the inherent limitations of a retrospective analysis. First, the primary outcome (opioid refills) may potentially be underestimated, as we captured this data based on clinical notes and orders in the electronic medical record system. It is possible that the need for opioid refills was missed in some patients who sought care outside of our health care system (and thus not recorded in our records). However, we extracted the data via a manual clinician review to optimize accuracy as much as possible. A prospective study would be needed to assess the incidence of postsurgical opioid refills more accurately. Second, an issue of generalizability is also of concern, as this is single-institution data. Model performance (eg, AUC) could decrease in a surgical population outside of this institution. To avoid the issue of overfitting and, thus, limited generalizability, we calculated the metrics from k-fold cross-validation and furthermore used a holdout data set for validation. What is needed is a high-quality prospective study that can more accurately capture the features and outcomes from each patient and, subsequently, be validated at external institutions.

Future Directions

Early identification of at-risk patients prior to their elective surgical procedure is the key. These patients can then be referred to establish care with a dedicated and comprehensive transitional pain program. Built on solid evidence-based medicine, this multidisciplinary transitional pain service includes anesthesiology, pharmacy, psychiatry, and physical therapy. Patients are often evaluated preoperatively to help manage expectations regarding anticipated postoperative pain and offer preoperative weaning when appropriate. This anesthesiologist-led team makes recommendations about intraoperative and immediate postoperative pain management [38], including predischarge and postdischarge tapering plans, if applicable. After the discharge, the transitional pain service can continue to manage these patients by using a multimodal approach with nonopioid medication, interventional procedures such as regional peripheral nerve blocks [39] or cryoanalgesia [40], as well as provide necessary psychological support [41]. Transitional pain clinics have been shown to reduce opioid use postoperatively, symptoms of anxiety and depression, pain catastrophizing, and pain [7,8]. Early identification of these clinical predictors, in conjunction with knowing the typical pain trajectories and patterns of common surgical procedures [42], can serve as the foundation for the basis of prescribing the right regimen and duration for the opioid prescription. Anesthesiology as a specialty, and especially in the setting of a dedicated acute pain service, is well positioned to take the lead in defining personalized pain medicine through all 3 phases of perioperative care [43].

Conclusions

Applying machine learning algorithms to electronic health data allows providers to develop models to predict more accurately and therefore appropriately allocate the limited health care resources (ie, transitional pain clinics). In this study, we showed that the need for regional anesthesia, high intraoperative opioid consumption, increased PACU pain scores, and opioid consumption were important features in models predicting outpatient opioid refills. Although providers are aware of the potential risk factors of opioid misuse, it remains challenging to accurately predict patients that will benefit from services as an outpatient. This prediction model serves as an example of a model that could be formalized into clinical decision support tools to help us better understand which patients will benefit from transitional pain clinics following ambulatory surgery.

Conflicts of Interest

SS is the Chief Executive Officer and founder of BrilliantBiome. RAG's institution has received funding and products for other research projects from Epimed International, Infutronics, and SPR Therapeutics for the following authors: RAG and ETS. The University of California San Diego is a consultant for Avanos, which is represented by RAG. This work was funded by the Division of Perioperative Informatics at the University of California San Diego.

References

1. Ladha KS, Neuman MD, Broms G, Bethell J, Bateman BT, Wijeysundera DN, et al. Opioid Prescribing After Surgery in the United States, Canada, and Sweden. JAMA Netw Open 2019 Sep 04;2(9):e1910734 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.10734] [Medline: 31483475]
2. Florence C, Luo F, Rice K. The economic burden of opioid use disorder and fatal opioid overdose in the United States, 2017. Drug Alcohol Depend 2021 Jan 01;218:108350 [FREE Full text] [doi: 10.1016/j.drugalcdep.2020.108350] [Medline: 33121867]
3. Florence CS, Zhou C, Luo F, Xu L. The Economic Burden of Prescription Opioid Overdose, Abuse, and Dependence in the United States, 2013. Med Care 2016 Oct;54(10):901-906 [FREE Full text] [doi: 10.1097/MLR.0000000000000625] [Medline: 27623005]

XSL•FO
RenderX

4. Brummett CM, Waljee JF, Goesling J, Moser S, Lin P, Englesbe MJ, et al. New Persistent Opioid Use After Minor and Major Surgical Procedures in US Adults. JAMA Surg 2017 Jun 21;152(6):e170504 [FREE Full text] [doi: 10.1001/jamasurg.2017.0504] [Medline: 28403427]
5. Clarke H, Soneji N, Ko DT, Yun L, Wijeysundera DN. Rates and risk factors for prolonged opioid use after major surgery: population based cohort study. BMJ 2014 Feb 11;348:g1251 [FREE Full text] [doi: 10.1136/bmj.g1251] [Medline: 24519537]
6. Sun EC, Mariano ER, Narouze S, Gabriel RA, Elsharkawy H, Gulur P, et al. Making a business plan for starting a transitional pain service within the US healthcare system. Reg Anesth Pain Med 2021 Aug;46(8):727-731. [doi: 10.1136/rapm-2021-102669] [Medline: 33879540]
7. Katz J, Weinrib A, Fashler SR, Katznelzon R, Shah BR, Ladak SS, et al. The Toronto General Hospital Transitional Pain Service: development and implementation of a multidisciplinary program to prevent chronic postsurgical pain. J Pain Res 2015;8:695-702 [FREE Full text] [doi: 10.2147/JPR.S91924] [Medline: 26508886]
8. Katz J, Weinrib AZ, Clarke H. Chronic postsurgical pain: From risk factor identification to multidisciplinary management at the Toronto General Hospital Transitional Pain Service. Can J Pain 2019;3(2):49-58 [FREE Full text] [doi: 10.1080/24740527.2019.1574537] [Medline: 35005419]
9. Featherall J, Anderson JT, Anderson LA, Bayless K, Anderson Z, Brooke BS, et al. A Multidisciplinary Transitional Pain Management Program Is Associated With Reduced Opioid Dependence After Primary Total Joint Arthroplasty. J Arthroplasty 2022 Jun;37(6):1048-1053. [doi: 10.1016/j.arth.2022.02.032] [Medline: 35181448]
10. Nair AA, Velagapudi MA, Lang JA, Behara L, Venigandla R, Velagapudi N, et al. Machine learning approach to predict postoperative opioid requirements in ambulatory surgery patients. PLoS One 2020;15(7):e0236833 [FREE Full text] [doi: 10.1371/journal.pone.0236833] [Medline: 32735604]
11. Lee S, Reid A, Tong S, Silveira L, Thomas JJ, Masaracchia MM. A Retrospective Review of Opioid Prescribing Practices for At-Risk Pediatric Populations Undergoing Ambulatory Surgery. J Pediatr Pharmacol Ther 2022;27(1):51-56 [FREE Full text] [doi: 10.5863/1551-6776-27.1.51] [Medline: 35002559]
12. Shanahan CW, Reding O, Holmdahl I, Keosaian J, Xuan Z, McAneny D, et al. Opioid analgesic use after ambulatory surgery: a descriptive prospective cohort study of factors associated with quantities prescribed and consumed. BMJ Open 2021 Aug 12;11(8):e047928 [FREE Full text] [doi: 10.1136/bmjopen-2020-047928] [Medline: 34385249]
13. Sridharan M, Samade R, J Kopechek K, Roebke AJ, Goyal KS, L Jones G, et al. The Effect of Patient and Surgical Factors on Opioid Prescription Requests Following Arthroscopic Rotator Cuff Repair. Arthrosc Sports Med Rehabil 2021 Jun;3(3):e707-e713 [FREE Full text] [doi: 10.1016/j.asmr.2021.01.008] [Medline: 34195635]
14. Gabriel RA, Harjai B, Prasad RS, Simpson S, Chu I, Fisch KM, et al. Machine learning approach to predicting persistent opioid use following lower extremity joint arthroplasty. Reg Anesth Pain Med 2022 May;47(5):313-319 [FREE Full text] [doi: 10.1136/rapm-2021-103299] [Medline: 35115414]
15. Clarke C, McClure A, Allen L, Hartford L, Van Koughnett JA, Gray D, et al. Opioid use after outpatient elective general surgery: quantifying the burden of persistent use. Br J Pain 2022 Aug;16(4):361-369. [doi: 10.1177/20494637211032907] [Medline: 36032343]
16. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. jair 2002 Jun 01;16:321-357. [doi: 10.1613/jair.953]
17. Anderson AB, Grazal CF, Balazs GC, Potter BK, Dickens JF, Forsberg JA. Can Predictive Modeling Tools Identify Patients at High Risk of Prolonged Opioid Use After ACL Reconstruction? Clin Orthop Relat Res 2020 Jul;478(7):-1618 [FREE Full text] [doi: 10.1097/CORR.0000000000001251] [Medline: 32282466]
18. Connor CW. Artificial Intelligence and Machine Learning in Anesthesiology. Anesthesiology 2019 Dec;131(6):1346-1359 [FREE Full text] [doi: 10.1097/ALN.0000000000002694] [Medline: 30973516]
19. Sun EC, Darnall BD, Baker LC, Mackey S. Incidence of and Risk Factors for Chronic Opioid Use Among Opioid-Naive Patients in the Postoperative Period. JAMA Intern Med 2016 Sep 01;176(9):1286-1293 [FREE Full text] [doi: 10.1001/jamainternmed.2016.3298] [Medline: 27400458]
20. Hooten WM, St Sauver JL, McGree ME, Jacobson DJ, Warner DO. Incidence and Risk Factors for Progression From Short-term to Episodic or Long-term Opioid Prescribing: A Population-Based Study. Mayo Clin Proc 2015 Jul;90(7):850-856 [FREE Full text] [doi: 10.1016/j.mayocp.2015.04.012] [Medline: 26141327]
21. Gabriel RA, Harjai B, Prasad RS, Simpson S, Chu I, Fisch KM, et al. Machine learning approach to predicting persistent opioid use following lower extremity joint arthroplasty. Reg Anesth Pain Med 2022 May;47(5):313-319 [FREE Full text] [doi: 10.1136/rapm-2021-103299] [Medline: 35115414]
22. Kim SC, Choudhry N, Franklin JM, Bykov K, Eikermann M, Lii J, et al. Patterns and predictors of persistent opioid use following hip or knee arthroplasty. Osteoarthritis Cartilage 2017 Sep;25(9):1399-1406 [FREE Full text] [doi: 10.1016/j.joca.2017.04.002] [Medline: 28433815]
23. Westermann RW, Mather RC, Bedard NA, Anthony CA, Glass NA, Lynch TS, et al. Prescription Opioid Use Before and After Hip Arthroscopy: A Caution to Prescribers. Arthroscopy 2019 Feb;35(2):453-460. [doi: 10.1016/j.arthro.2018.08.056] [Medline: 30612773]

XSL•FO
RenderX

24. Rao AG, Chan PH, Prentice HA, Paxton EW, Navarro RA, Dillon MT, et al. Risk factors for postoperative opioid use after elective shoulder arthroplasty. J Shoulder Elbow Surg 2018 Nov;27(11):1960-1968. [doi: 10.1016/j.jse.2018.04.018] [Medline: 29891412]
25. Inacio MCS, Hansen C, Pratt NL, Graves SE, Roughead EE. Risk factors for persistent and new chronic opioid use in patients undergoing total hip arthroplasty: a retrospective cohort study. BMJ Open 2016 Apr 29;6(4):e010664 [FREE Full text] [doi: 10.1136/bmjopen-2015-010664] [Medline: 27130165]
26. Lawal OD, Gold J, Murthy A, Ruchi R, Bavry E, Hume AL, et al. Rate and Risk Factors Associated With Prolonged Opioid Use After Surgery: A Systematic Review and Meta-analysis. JAMA Netw Open 2020 Jun 01;3(6):e207367 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.7367] [Medline: 32584407]
27. Clarke HA, Manoo V, Pearsall EA, Goel A, Feinberg A, Weinrib A, et al. Consensus Statement for the Prescription of Pain Medication at Discharge after Elective Adult Surgery. Can J Pain 2020 Mar 08;4(1):67-85 [FREE Full text] [doi: 10.1080/24740527.2020.1724775] [Medline: 33987487]
28. Hill MV, McMahon ML, Stucke RS, Barth RJ. Wide Variation and Excessive Dosage of Opioid Prescriptions for Common General Surgical Procedures. Ann Surg 2017 Apr;265(4):709-714. [doi: 10.1097/SLA.0000000000001993] [Medline: 27631771]
29. Eid AI, DePesa C, Nordestgaard AT, Kongkaewpaisan N, Lee JM, Kongwibulwut M, et al. Variation of Opioid Prescribing Patterns among Patients undergoing Similar Surgery on the Same Acute Care Surgery Service of the Same Institution: Time for Standardization? Surgery 2018 Nov;164(5):926-930. [doi: 10.1016/j.surg.2018.05.047] [Medline: 30049481]
30. Wunsch H, Wijeysundera DN, Passarella MA, Neuman MD. Opioids Prescribed After Low-Risk Surgical Procedures in the United States, 2004-2012. JAMA 2016 Apr 19;315(15):1654-1657 [FREE Full text] [doi: 10.1001/jama.2016.0130] [Medline: 26978756]
31. Goesling J, Moser SE, Zaidi B, Hassett AL, Hilliard P, Hallstrom B, et al. Trends and predictors of opioid use after total knee and total hip arthroplasty. Pain 2016 Jun;157(6):1259-1265 [FREE Full text] [doi: 10.1097/j.pain.0000000000000516] [Medline: 26871536]
32. Bicket MC, Long JJ, Pronovost PJ, Alexander GC, Wu CL. Prescription Opioid Analgesics Commonly Unused After Surgery: A Systematic Review. JAMA Surg 2017 Nov 01;152(11):1066-1071 [FREE Full text] [doi: 10.1001/jamasurg.2017.0831] [Medline: 28768328]
33. Compton WM, Jones CM, Baldwin GT. Relationship between Nonmedical Prescription-Opioid Use and Heroin Use. N Engl J Med 2016 Jan 14;374(2):154-163. [doi: 10.1056/NEJMra1508490] [Medline: 26760086]
34. Bartels K, Mayes LM, Dingmann C, Bullard KJ, Hopfer CJ, Binswanger IA. Opioid Use and Storage Patterns by Patients after Hospital Discharge following Surgery. PLoS One 2016;11(1):e0147972 [FREE Full text] [doi: 10.1371/journal.pone.0147972] [Medline: 26824844]
35. Howard R, Brown CS, Lai Y, Gunaseelan V, Chua K, Brummett C, et al. The Association of Postoperative Opioid Prescriptions with Patient Outcomes. Ann Surg 2022 Dec 01;276(6):e1076-e1082. [doi: 10.1097/SLA.0000000000004965] [Medline: 34091508]
36. Porter ED, Bessen SY, Molloy IB, Kelly JL, Ramkumar N, Phillips JD, et al. Guidelines for Patient-CenteredOpioid Prescribing and Optimal FDA-Compliant Disposal of Excess Pills after Inpatient Operation: Prospective Clinical Trial. J Am Coll Surg 2021 Jun;232(6):823-835.e2. [doi: 10.1016/j.jamcollsurg.2020.12.057] [Medline: 33640521]
37. Gupta S, Mohta A, Gottumukkala V. Opioid-free anesthesia-caution for a one-size-fits-all approach. Perioper Med (Lond) 2020;9:16 [FREE Full text] [doi: 10.1186/s13741-020-00147-3] [Medline: 32566148]
38. Huang A, Azam A, Segal S, Pivovarov K, Katznelson G, Ladak SS, et al. Chronic postsurgical pain and persistent opioid use following surgery: the need for a transitional pain service. Pain Manag 2016 Oct;6(5):435-443 [FREE Full text] [doi: 10.2217/pmt-2016-0004] [Medline: 27381204]
39. Chitnis SS, Tang R, Mariano ER. The role of regional analgesia in personalized postoperative pain management. Korean J Anesthesiol 2020 Oct;73(5):363-371 [FREE Full text] [doi: 10.4097/kja.20323] [Medline: 32752602]
40. Ilfeld BM, Finneran JJ. Cryoneurolysis and Percutaneous Peripheral Nerve Stimulation to Treat Acute Pain. Anesthesiology 2020 Nov 01;133(5):1127-1149 [FREE Full text] [doi: 10.1097/ALN.0000000000003532] [Medline: 32898231]
41. Doan LV, Blitz J. Preoperative Assessment and Management of Patients with Pain and Anxiety Disorders. Curr Anesthesiol Rep 2020;10(1):28-34 [FREE Full text] [doi: 10.1007/s40140-020-00367-9] [Medline: 32435161]
42. Mariano ER, El-Boghdadly K, Ilfeld BM. Using postoperative pain trajectories to define the role of regional analgesia in personalised pain medicine. Anaesthesia 2021 Feb;76(2):165-169 [FREE Full text] [doi: 10.1111/anae.15067] [Medline: 32368794]
43. Larach DB, Hah JM, Brummett CM. Perioperative Opioids, the Opioid Crisis, and the Anesthesiologist. Anesthesiology 2022 Apr 01;136(4):594-608 [FREE Full text] [doi: 10.1097/ALN.0000000000004109] [Medline: 35108351]

Abbreviations

AUC: area under the curve
MEQ: morphine equivalents

XSL•FO
RenderX

OR: odds ratio
PACU: postanesthesia care unit
SHAP: Shapley Additive Explanations
SMOTE: Synthetic Minority Oversampling Technique

Edited by T Leung; submitted 21.06.22; peer-reviewed by H Lee, K McKelvey, S Liu; comments to author 23.11.22; revised version received 06.12.22; accepted 24.01.23; published 08.02.23.

Please cite as:
Gabriel RA, Simpson S, Zhong W, Burton BN, Mehdipour S, Said ET
A Neural Network Model Using Pain Score Patterns to Predict the Need for Outpatient Opioid Refills Following Ambulatory Surgery: Algorithm Development and Validation
JMIR Perioper Med 2023;6:e40455
URL: https://periop.jmir.org/2023/1/e40455
doi:10.2196/40455
PMID:36753316

©Rodney Allanigue Gabriel, Sierra Simpson, William Zhong, Brittany Nicole Burton, Soraya Mehdipour, Engy Tadros Said. Originally published in JMIR Perioperative Medicine (http://periop.jmir.org), 08.02.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Perioperative Medicine, is properly cited. The complete bibliographic information, a link to the original publication on http://periop.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Association Between Borderline Dysnatremia and Perioperative Morbidity and Mortality: Retrospective Cohort Study of the American College of Surgeons National Surgical Quality Improvement Program Database

Jacob H Cole[1,2,3], MD; Krista B Highland[2], PhD; Scott B Hughey[3,4], MD; Brendan J O'Shea[4], MD; Thomas Hauert[4], DO; Ashton H Goldman[5], MD; George C Balazs[5], MD; Gregory J Booth[2,3,4], MD

[1]Department of Anesthesiology, Perioperative, and Pain Medicine, Brigham and Women's Hospital, Boston, MA, United States
[2]Department of Anesthesiology, Uniformed Services University, Bethesda, MD, United States
[3]Naval Biotechnology Group, Naval Medical Center Portsmouth, Portsmouth, VA, United States
[4]Department of Anesthesiology and Pain Medicine, Naval Medical Center Portsmouth, Portsmouth, VA, United States
[5]Department of Orthopedic Surgery, Naval Medical Center Portsmouth, Portsmouth, VA, United States

Corresponding Author:
Jacob H Cole, MD
Department of Anesthesiology, Perioperative, and Pain Medicine
Brigham and Women's Hospital
75 Francis Street
Boston, MA, 02115
United States
Phone: 1 5702627798
Email: cole.jacob.h@gmail.com

Abstract

Background: Hyponatremia and hypernatremia, as conventionally defined (<135 mEq/L and >145 mEq/L, respectively), are associated with increased perioperative morbidity and mortality. However, the effects of subtle deviations in serum sodium concentration within the normal range are not well-characterized.

Objective: The purpose of this analysis is to determine the association between borderline hyponatremia (135-137 mEq/L) and hypernatremia (143-145 mEq/L) on perioperative morbidity and mortality.

Methods: A retrospective cohort study was performed using data from the American College of Surgeons National Surgical Quality Improvement Program database. This database is a repository of surgical outcome data collected from over 600 hospitals across the United States. The National Surgical Quality Improvement Program database was queried to extract all patients undergoing elective, noncardiac surgery from 2015 to 2019. The primary predictor variable was preoperative serum sodium concentration, measured less than 5 days before the index surgery. The 2 primary outcomes were the odds of morbidity and mortality occurring within 30 days of surgery. The risk of both outcomes in relation to preoperative serum sodium concentration was modeled using weighted generalized additive models to minimize the effect of selection bias while controlling for covariates.

Results: In the overall cohort, 1,003,956 of 4,551,726 available patients had a serum sodium concentration drawn within 5 days of their index surgery. The odds of morbidity and mortality across sodium levels of 130-150 mEq/L relative to a sodium level of 140 mEq/L followed a nonnormally distributed U-shaped curve. The mean serum sodium concentration in the study population was 139 mEq/L. All continuous covariates were significantly associated with both morbidity and mortality (P<.001). Preoperative serum sodium concentrations of less than 139 mEq/L and those greater than 144 mEq/L were independently associated with increased morbidity probabilities. Serum sodium concentrations of less than 138 mEq/L and those greater than 142 mEq/L were associated with increased mortality probabilities. Hypernatremia was associated with higher odds of both morbidity and mortality than corresponding degrees of hyponatremia.

Conclusions: Among patients undergoing elective, noncardiac surgery, this retrospective analysis found that preoperative serum sodium levels less than 138 mEq/L and those greater than 142 mEq/L are associated with increased morbidity and mortality, even within currently accepted “normal” ranges. The retrospective nature of this investigation limits the ability to make causal determinations for these findings. Given the U-shaped distribution of risk, past investigations that assume a linear relationship

XSL•FO
RenderX

between serum sodium concentration and surgical outcomes may need to be revisited. Likewise, these results question the current definition of perioperative eunatremia, which may require future prospective investigations.

(JMIR Perioper Med 2023;6:e38462) doi:10.2196/38462

KEYWORDS

hypernatremia; hyponatremia; perioperative care; postoperative complications; reference values; sodium; morbidity; mortality; database; data; cohort; surgery; sodium; preoperative; serum

Introduction

Abnormal preoperative sodium levels are associated with multiple adverse outcomes, including increased risk of venous thromboembolism, major bleeding and return to the operating room, perioperative coronary events, wound infection, and prolonged postoperative length of hospital stay [1-6]. Both hyponatremia and hypernatremia are associated with an increased risk of perioperative mortality [2,4,5]. Past investigations in nonsurgical populations suggest that optimizing sodium intake may reduce the risk of mortality [7,8]. While these studies provide a clinical rationale for intervention in the presence of hyponatremia or hypernatremia, the granularity of results has been limited due to broad categorizations of hyponatremia and hypernatremia.

Many previous studies investigating patient outcomes categorize sodium levels as hyponatremic (serum sodium concentration less than 135 mEq/L), eunatremic, and hypernatremic (serum sodium concentration greater than 145 mEq/L) [1,9-11]. Some studies also identified an increased risk of in-hospital and 1-year mortality in hospitalized patients with mild hyponatremia (125-134 mEq/L) and hypernatremia (146-150 mEq/L) [12,13]. Such evidence indicates that there are gradations of risk per sodium level *outside* of the eunatremic range, but it is unknown if such gradations of risk occur *within* the eunatremic range. Therefore, a more granular resolution is needed to determine if there is an increased risk of poor postoperative outcomes in patients within the range of serum sodium concentrations that are currently accepted as normal.

The culmination of research to date indicates that the role of sodium in morbidity and mortality risk is broad across a variety of surgeries, including hip arthroplasty [7,8], lower extremity arthroplasty [14], cervical spinal fusion [15], and cardiac surgery [9,16]. Moreover, risk prediction models, including those based on the American College of Surgeons National Surgical Quality Improvement Program (ACS NSQIP) data, indicate that sodium level, when categorized (eg, hyponatremia, eunatremic, and hypernatremia), is an important indicator of postsurgical morbidity and mortality in a large surgically diverse sample [17]. Such risk models do not allow clinicians to delineate an ideal target for clinical intervention. Taken together, there is a need to provide clinically informative research that evaluates the nonnormally distributed relationship between sodium levels, morbidity, and mortality across a large surgical population. Therefore, the purpose of this investigation was to explore the potential nonlinear relationship between preoperative sodium levels, modeled as a continuous predictor, and the odds of 30-day postoperative morbidity and mortality in patients undergoing elective, noncardiac surgery. We hypothesized that preoperative serum sodium concentration was independently associated with increased odds of both postoperative morbidity and mortality when modeled as a continuous variable, assuming a reference normal serum sodium concentration of 140 mEq/L.

Methods

Ethical Considerations

This study is a retrospective cohort design and was approved by the Naval Medical Center Portsmouth's Institutional Review Board (NMCP.2021.0054).

Study Design and Data Source

Data from the ACS NSQIP database during the years 2015-2019 were obtained. These data come from over 700 hospitals and are collected using well-described methods to assure a high level of validity [18]. Noncardiac surgical procedures were included using current procedural terminology (CPT) codes 10000-32999 and 34000-69999. Patients undergoing cardiac surgery were excluded from this analysis due to the unique risks associated with that patient population, including the risks associated with cardiopulmonary bypass. Similar to previous investigations [19], we excluded minor surgeries such as endoscopies (CPT 43200-43272, 45300-45392, 46600-46608) and minor musculoskeletal procedures (CPT 29000-29750). Additionally, patients were excluded if they underwent emergency surgery.

The following demographic and health data were collected for each patient: CPT code, age, race, ethnicity, height, weight, sex assigned in the medical record, functional status, American Society of Anesthesiologists (ASA) Physical Score, sodium level, hematocrit, creatinine, steroid use, ascites, sepsis or septic shock, ventilator dependence, disseminated cancer, diabetes, hypertension, weight loss (at least 10% in the past year), congestive heart failure (CHF), dyspnea, smoking, chronic obstructive pulmonary disease, and dialysis. Patient records were included based on the following criteria: sodium, hematocrit, and creatinine assessment <5 days prior to surgery; BMI of >12 and <60; ages 18 to 89 years; hematocrit of >21% and <50%; sodium level of ≥130 mEq/L and ≤150 mEq/L; creatinine level of ≥0.5 mg/dL and ≤4.0 mg/dL; and undergoing surgery under a primary CPT listed in at least 50 patient records.

Exposure

The primary exposure was the preoperative sodium level. *A priori*, the serum sodium level of 140 mEq/L was empirically determined to be the reference value for the development of statistical models.

XSL•FO
RenderX

Outcomes

The 2 primary outcomes were defined as aggregate morbidity within 30 days of index surgery and mortality within 30 days of index surgery. Aggregate morbidity included any of the following: cardiac arrest, myocardial infarction, cerebrovascular accident, deep vein thrombosis, pulmonary embolism, postoperative sepsis or septic shock, renal insufficiency or failure, reintubation, failure to wean from the ventilator, pneumonia, wound dehiscence, or surgical site infection (including superficial, deep, or organ space). Details regarding the standardized definitions of these variables have been previously published [19].

Statistical Analysis

Univariate and Bivariate Analyses

First, nonparametric analyses (eg, chi-square, Kruskal-Wallis rank sum, and Mann-Whitney U tests) examined differences between patient records that were and were not included in the analyses. Next, bivariate analyses evaluated differences in demographic characteristics and medical comorbidities by morbidity and mortality status. Bivariate analyses were performed using the *TableOne* R package (R Foundation) [20]. Due to the elevated likelihood of rejecting the null hypothesis ($P<.05$) in large samples and because the information rendered by the P value does not describe the strength of differences, both the P value and the standardized mean difference are reported for bivariate analyses. Standardized mean difference is reported specifically to describe the effect size of the included demographic characteristics and medical comorbidities on the outcomes of morbidity and mortality.

Inverse Probability Weights

Given the potential for selection bias in this analysis, outcome models included weights corresponding to the inverse probability of meeting inclusion criteria. This previously validated method accounts for selection bias due to missing predictor data [21]. Inverse probability weights were constructed through a multistep process. First, a generalized additive model (GAM) was conducted using the *mgcv* R package [22] to estimate the propensity of record inclusion. GAMs allowed for the modeling of nonlinear relationships between continuous predictors and the outcomes (smooth effects). In the GAM, the binary outcome was recorded as exclusion (0) versus inclusion (1), and the predictors were covariates associated with included versus excluded status. Sodium, creatinine, and hematocrit were not used in this analysis, as the lack of preoperative laboratory data was indicative of an excluded status. To account for the role of primary CPT in the propensity to be included, the proportion (%) of included patients per primary CPT was calculated. This proportion was included in the GAM as an additional covariate. The predicted and fitted values indicated the propensity of record inclusion given demographic characteristics, medical comorbidities, and primary CPT. Lastly, the propensity scores were transformed into inverse probability weights through the following formula: Inverse probability weight = (Included status / Propensity score) + ((1 Included status) / (1 Propensity score)). These weights were used to control for potential selection bias in subsequent outcome models [23].

Generalized Additive Models

The previously described factors associated with morbidity and mortality within the NSQIP database were included as covariates in 2 separate GAMs. One model was generated to predict aggregate morbidity, and the other to predict mortality. If missing data in the included sample was >1%, multiple imputations were planned. To assess the degree of multicollinearity, the *performance* R package was used to compute the variance inflation factor of each fixed covariate; a variance inflation factor <5 indicated acceptable levels of multicollinearity. GAM results were extracted using the *sjPlot* R package [24]. Estimated conditional means (95% CI) were calculated using the *ggeffects* R package [25]. Both the adjusted odds ratios (95% CI) and adjusted relative risks (RRs, 95% CI) of morbidity and mortality at sodium levels 130-150 mEq/L, relative to the *a priori* defined reference of 140 mEq/L, were calculated as well. The *ggplot2* [26] and *ggpubr* [27] R packages were used to construct customized plots of model results. Statistical significance was indicated by $P<.05$.

Sensitivity Analysis

Sensitivity analyses were performed using E-values [28] and stratification of the included sample by the previously calculated propensity scores. The *EValue* R package [29] was used to calculate E-values corresponding to each RR of sodium levels 130-150 mEq/L. E-values indicate the strength a confounding variable would need to have on both the predictor (sodium) and outcome, beyond the effects of covariates already included in the model, to render the effect of sodium on the outcome null [30]. As such, E-values provide an assumption-free means of evaluating the robustness of model results [28]. For comparison purposes, the RR (95% CI) of fixed effects was also calculated. Within the included sample, propensity scores corresponding to the propensity to be included in analyses were divided into terciles. The outcome GAMs were replicated without the weights in the subsample of included records with the lowest tercile of propensity scores. Sensitivity analyses were graphically rendered for comparison purposes.

Results

Sample Description

Of the 4,551,726 patient records available, 1,003,956 met all inclusion criteria. Most patient records were excluded due to laboratory assessments occurring more than 4 days from surgery or not at all (n=3,388,178), continuous variables outside of the prespecified ranges (n=145,458), and a primary CPT that was not represented in at least 50 patient records (n=14,134). Bivariate analyses indicated that those included versus excluded differed across all identified demographic characteristics and medical comorbidities (Multimedia Appendix 1). In the included sample, 15,474 (0.3%) patient records had missing data; therefore, no imputation was performed. Morbidity and mortality rates in the included cohort were 8.5% and 1.3%, respectively. Descriptive statistics are reported (Table 1). Morbidity (Table 2) and mortality status (Table 3) are also reported. Bivariate

XSL•FO
RenderX

test results indicated that all demographic characteristics and medical comorbidities were associated with morbidity and mortality status. As such, all of these factors were included as covariates in the GAMs.

Table 1. Descriptive statistics of the overall sample (N=977,343).

Characteristics	Overall
Age (years), median (IQR)	60.0 (46.0-71.0)
Sex, n (%)	
Male	452,054 (45.0)
Female	551,884 (55.0)
Race and ethnicity, n (%)	
White	654,377 (65.2)
American Indian and Alaska Native	6366 (0.6)
Asian	27,927 (2.8)
Black	111,166 (11.1)
Latino	73,748 (7.3)
Native Hawaiian and Pacific Islander	3575 (0.4)
Other	2451 (0.2)
Unknown	124,346 (12.4)
BMI, median (IQR)	28.66 (24.69-33.67)
ASA[a] physical status, n (%)	
I	56,585 (5.7)
II	387,503 (38.7)
III	477,321 (47.7)
IV	79,712 (8.0)
Presence of comorbidities, n (%)	
Hypertension	481,752 (48.0)
Diabetes	191,078 (19.6)
COPD[b]	56,487 (5.6)
History of smoking	200,591 (20.0)
Chronic steroid use	48,421 (4.8)
Congestive heart failure	14,385 (1.4)
Active cancer diagnosis	40,880 (4.1)
Sepsis or septic shock	37,231 (3.8)
Preoperative laboratory values, median (IQR)	
Sodium (mEq/L)	139 (137-141)
Hematocrit (%)	39.2 (35.2-42.5)
Creatinine (mg/dL)	0.84 (0.70-1.01)
Percent CPT[c] morbidity (IQR)	5.30 (2.63-11.28)
Percent CPT mortality (IQR)	0.24 (0.08-1.47)

[a]ASA: American Society of Anesthesiologists.

[b]COPD: chronic obstructive pulmonary disease.

[c]CPT: current procedural terminology.

XSL•FO
RenderX

Table 2. Aggregate morbidity outcomes status.

Characteristics	No morbidity (N=918,385)	Morbidity (N=85,571)	*P* value	SMD[a]
Age (years), median (IQR)	59.0 (45.0-70.0)	65.0 (54.0-75.0)	<.001	0.38
Sex, n (%)			<.001	0.08
Male	410,367 (44.7)	41,687 (48.7)		
Female	508,002 (55.3)	43,882 (51.3)		
Race and ethnicity, n (%)			<.001	0.1
White	597,599 (65.1)	56,778 (66.4)		
American Indian and Alaska Native	5814 (0.6)	552 (0.6)		
Asian	25,943 (2.8)	1984 (2.3)		
Black	100,829 (11.0)	10,337 (12.1)		
Latino	69,192 (7.5)	4556 (5.3)		
Native Hawaiian and Pacific Islander	3303 (0.4)	272 (0.3)		
Other	2281 (0.2)	170 (0.2)		
Unknown	113,424 (12.4)	10,922 (12.8)		
BMI, median (IQR)	28.69 (24.74-33.67)	28.69 (24.74-33.67)	<.001	0.03
ASA[b] physical status, n (%)			<.001	0.57
I	55,391 (6.0)	1194 (1.4)		
II	369,550 (40.3)	17,953 (21.1)		
III	426,405 (46.6)	50,916 (59.8)		
IV	64,588 (7.1)	15,124 (17.8)		
Presence of comorbidities, n (%)				
Hypertension	430,631 (46.9)	51,121 (59.7)	<.001	0.26
Diabetes	168,658 (18.4)	22,420 (26.2)	<.001	0.21
COPD[c]	47,235 (5.1)	9252 (10.8)	<.001	0.21
History of smoking	181,440 (19.8)	19,151 (22.4)	<.001	0.06
Chronic steroid use	41,464 (4.5)	6957 (8.1)	<.001	0.15
Congestive heart failure	11,224 (1.2)	3161 (3.7)	<.001	0.16
Active cancer diagnosis	33,660 (3.7)	7220 (8.4)	<.001	0.20
Sepsis or septic shock	31,324 (3.4)	5907 (6.9)	<.001	0.17
Preoperative laboratory values, median (IQR)				
Sodium (mEq/L)	139 (137-141)	139 (137-141)	<.001	0.12
Hematocrit (%)	39.4 (35.6-42.6)	37.0 (32.0-41.0)	<.001	0.41
Creatinine (m)g/dL	0.83 (0.70-1.00)	0.88 (0.70-1.11)	<.001	0.22
Percent CPT[d] morbidity (IQR)	5.02 (2.63-10.26)	12.38 (5.93-19.54)	<.001	0.81
Percent CPT mortality (IQR)	0.22 (0.08-1.15)	1.25 (0.33-3.05)	<.001	0.51

[a]SMD: standardized mean difference.

[b]ASA: American Society of Anesthesiologists.

[c]COPD: chronic obstructive pulmonary disease.

[d]CPT: current procedural terminology.

XSL•FO
RenderX

Table 3. Mortality outcome status.

Characteristics	No mortality (N=991,327)	Mortality (N=12,629)	*P* value	SMD[a]
Age (years), median (IQR)	60.00 (46.00-71.00)	75.00 (66.00-82.00)	<.001	1.02
Sex, n (%)			<.001	0.19
Male	445,213 (44.9)	6841 (54.2)		
Female	546,096 (55.1)	5788 (45.8)		
Race and ethnicity, n (%)			<.001	0.24
White	645,010 (65.1)	9367 (74.2)		
American Indian & Alaska Native	6315 (0.6)	51 (0.4)		
Asian	27,682 (2.8)	245 (1.9)		
Black	109,799 (11.1)	1367 (10.8)		
Latino	73,256 (7.4)	492 (3.9)		
Native Hawaiian and Pacific Islander	3546 (0.4)	29 (0.2)		
Other	2430 (0.2)	21 (0.2)		
Unknown	123,289 (12.4)	1057 (8.4)		
BMI, median (IQR)	28.69 (24.74-33.73)	25.99 (22.20-30.99)	<.001	0.35
ASA[b] physical status, n (%)			<.001	1.25
I	56,574 (5.7)	11 (0.1)		
II	386,890 (39.1)	613 (4.9)		
III	470,707 (47.6)	6614 (53.2)		
IV	74,518 (7.5)	5194 (41.8)		
Presence of comorbidities, n (%)				
Hypertension	472,832 (47.7)	8920 (70.6)	<.001	0.48
Diabetes	187,320 (18.9)	3758 (29.8)	<.001	0.29
COPD[c]	54,107 (5.5)	2380 (18.8)	<.001	0.42
History of smoking	198,133 (20.0)	2458 (19.5)	.15	0.01
Chronic steroid use	47,044 (4.7)	1377 (10.9)	<.001	0.23
Congestive heart failure	13,072 (1.3)	1313 (10.4)	<.001	0.39
Active cancer diagnosis	38,451 (3.9)	2429 (19.2)	<.001	0.50
Sepsis or septic shock	35,342 (3.6)	1889 (1.5)	<.001	0.42
Preoperative laboratory values, median (IQR)				
Sodium (mEq/L)	139 (137-141)	138 (136-141)	<.001	0.19
Hematocrit (%)	39.3 (35.3-42.5)	33.0 (28.8-38.0)	<.001	0.88
Creatinine (mg/dL)	0.83 (0.70-1.01)	1.00 (0.76-1.44)	<.001	0.54
Percent CPT[d] morbidity (IQR)	5.30 (2.63-10.90)	12.54 (9.08-18.46)	<.001	0.91
Percent CPT mortality (IQR)	0.23 (0.08-1.37)	2.91 (1.25-5.09)	<.001	1.05

[a]SMD: standardized mean difference.

[b]ASA: American Society of Anesthesiologists.

[c]COPD: chronic obstructive pulmonary disease.

[d]CPT: current procedural terminology.

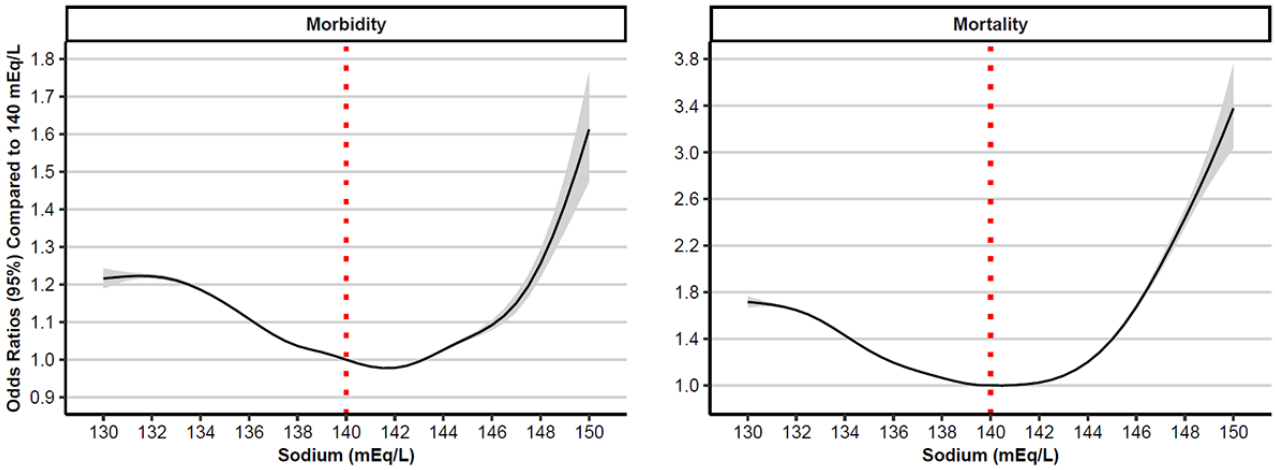
GAM Results

In both outcome GAMs, all continuous covariates (age, BMI, sodium, hematocrit, creatinine, and percent CPT morbidity or mortality) were modeled as smooth terms and were substantially associated with both morbidity and mortality. Across both models, patients assigned male in the medical record with an elevated ASA status, steroid use, sepsis or septic shock, cancer, a positive smoking status, chronic obstructive pulmonary

XSL•FO
RenderX

disease, renal failure, and CHF were more likely to experience morbidity and mortality compared to their reference counterparts (Multimedia Appendix 2). When controlling for other demographic characteristics and medical comorbidities, patients whose race and ethnicity were listed as Asian or Latino had a lower probability of morbidity and mortality relative to White patients. Similarly, White patients had a greater probability of morbidity relative to Native Hawaiian and Pacific Islander patients, but White patients had a lower probability of morbidity than patients of unknown race and ethnicity. Patients with diabetes and hypertension had a greater probability of morbidity relative to those without these conditions. The odds of morbidity and mortality across sodium levels of 130-150 mEq/L relative to a sodium level of 140 mEq/L followed a U-shaped curve (Figure 1).

Figure 1. Odds ratios (95% CI) of morbidity (left) and mortality (right).



Sensitivity Analyses

The E-values corresponding to the RR of morbidity and mortality across sodium levels are shown in Figures S1A and S1B in Multimedia Appendix 3. For example, at a sodium level of 135 mEq/L, the morbidity (RR 1.07, 95% CI 1.07-1.07) and mortality (RR 1.30, 95% CI 1.30-1.30) would be rendered null if an unmeasured confounder was associated with both sodium and the outcome by a RR of 1.35-fold (lower 95% CI 1.35) and 1.93-fold (lower 95% CI 1.92), respectively. For reference, these E-values are similar to the effects of ASA I versus III on morbidity (RR 1.33, 95% CI 1.32-1.33) and CHF on mortality (RR 1.97, 95% CI 1.94-1.99). The RR (95% CI) for fixed effects on both outcomes are reported in Multimedia Appendix 2.

Lastly, GAMs evaluating morbidity and mortality were conducted on a subsample of the included group with the lowest tercile propensity scores (n=333,701). Model results were similar to the main analysis, such that the effect of sodium was significant (P<.001), and the nonlinear pattern followed a U-shape (Figures S1C and S1D in Multimedia Appendix 3).

Discussion

Principal Results

This exploratory analysis calls into question the current understanding of the "normal" range of serum sodium levels (135-145 mEq/L) within the context of perioperative care, as values of serum sodium concentration within this range of normal values were associated with 30-day aggregate morbidity and mortality. By examining preoperative sodium levels in over 1 million patients as a continuous variable instead of the commonly used categories (eg, hyponatremic, eunatremic, and hypernatremic), this study provides improved granularity on the association between small deviations in sodium and perioperative outcomes. As such, what is considered "normal" sodium values in the general population may not be normal in patients undergoing elective noncardiac surgery.

Comparison With Prior Work

As health care shifts to value-based care, these findings may also play a role in evaluating value-based perioperative practices. For example, recent evidence using NSQIP data indicates that preoperative laboratory assessment is not associated with the odds of postoperative complications and readmission in patients undergoing ambulatory surgery with an ASA I or II status, thereby suggesting the low value of preoperative laboratory assessment [31]. However, such findings may be premised on clinician practices that are contingent on a definition of "normal" that is, per these findings, associated with increased risk of aggregate morbidity and mortality (eg, ~135 mEq/L). Given the potential impact of these findings, combined with the lack of causal assumptions that can be made, future work is needed to assess whether clinical intervention addressing high- and low-normal sodium serum concentrations improves clinical outcomes and value-based care.

Strengths and Limitations

This study possessed several strengths. Though the main variable of interest was serum sodium concentrations, models included many demographic characteristics and medical comorbidities that have previously been shown to be substantially associated with aggregate morbidity and mortality risk. These factors included other laboratory values (eg, creatinine and hematocrit) that may also warrant further inspection, given their relationship with postoperative outcomes. By controlling for these covariates and using a weighted approach based on the inverse probability of record inclusion, the results of this study are likely

XSL•FO
RenderX

generalizable to adult patients undergoing any elective, noncardiac surgery in the United States. We restricted records to those with laboratory results collected less than 5 days before surgery, thereby increasing the likelihood that the recorded values actually reflected serum sodium levels at the time of surgery.

This study was tempered by several limitations. First, no causal conclusions can be drawn from the study due to the retrospective, associative nature of the study design and analytic approach. Additionally, there may be several covariates, including specific health conditions, medication receipt (both in the days leading up to surgery and perioperatively), preoperative recommendations (eg, fasting), and prior health care received, that are neither collected in the NSQIP database nor included in the analysis but could be associated with morbidity and mortality. While this database is a robust and extensive collection of surgical outcome data in the United States [18], the inclusion of other covariates mentioned above could serve to refine this model and provide more specific areas of research to explore. Examples of other potential confounders include medications, preoperative fasting, and certain comorbidities, which themselves may be associated with abnormal sodium levels. When considering the potential impact of missing confounders on model results, E-values indicated that any confounder would need to surpass the strength of most fixed covariates within our models and account for unique variance not otherwise accounted for by current covariates to render the effect of sodium null.

Conclusions

This analysis indicated that both preoperative hyponatremia and preoperative hypernatremia were associated with an increased risk of 30-day aggregate morbidity and mortality. The relationship was nonlinear, such that the risk increased with further deviation from a serum sodium concentration of 140. While prior investigations have demonstrated that dysnatremia is a modifiable risk factor and optimization of preoperative serum sodium levels may represent an opportunity for a reduction in both perioperative morbidity and mortality [31], this study suggests that preoperative serum sodium levels that are within the currently accepted upper and lower limits of normal are likely indicative of elevated risk. As such, future prospective studies are needed to better confer sodium level ranges associated with optimized outcomes after surgery, as well as the potential to directly alter patients' serum sodium concentrations to improve postoperative outcomes.

Disclaimer

The views expressed are solely those of the authors and do not reflect the official policy or position of the Uniformed Services University, US Army, US Navy, US Air Force, the Department of Defense, the US Government, or the Henry M Jackson Foundation for the Advancement of Military Medicine, Inc. We are military service members. This work was prepared as part of our official duties. Title 17 U.S.C. 105 provides that "Copyright protection under this title is not available for any work of the United States Government." Title 17 U.S.C. 101 defines a United States Government work as a work prepared by a military service member or employee of the United States Government as part of that person's official duties. The American College of Surgeons National Surgical Quality Improvement Program (ACS NSQIP) and the hospitals participating in the ACS NSQIP are the sources of the data used herein; they have not verified and are not responsible for the statistical validity of the data analysis or the conclusions derived by the authors.

Authors' Contributions

JHC, SBH, and GJB were responsible for the initial conceptualization of this investigation. JHC, SBH, AHG, GCB, and GJB developed the formal methodology that was used in this study. JHC, BJO, TH, and GJB all participated in the background investigation for this work. KBH was instrumental in the software development, formal analysis, and data curation of this investigation. JHC, SBH, BJO, TH, and GJB all assisted in the creation of the original draft, while JHC, KBH, SBH, AHG, GCB, and GJB all facilitated the review and editing process.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Bivariate differences between excluded versus included records.
[XLSX File (Microsoft Excel File), 13 KB - periop_v6i1e38462_app1.xlsx]

Multimedia Appendix 2
Risk ratios (95% CI) of fixed covariates across outcomes.
[XLSX File (Microsoft Excel File), 10 KB - periop_v6i1e38462_app2.xlsx]

Multimedia Appendix 3

XSL•FO
RenderX

Sensitivity analyses of sodium effects. (A,B) Adjusted relative risk (95%CI) and E-values (lower 95% CI) of morbidity and mortality, respectively. (C,D) Odds ratios (95%CI) of morbidity and mortality, respectively.
[PDF File (Adobe PDF File), 9 KB - periop_v6i1e38462_app3.pdf]

References

1. Temraz S, Tamim H, Mailhac A, Taher A. Could sodium imbalances predispose to postoperative venous thromboembolism? An analysis of the NSQIP database. Thromb J 2018;16:11 [FREE Full text] [doi: 10.1186/s12959-018-0165-5] [Medline: 29988709]
2. Leung AA, McAlister FA, Rogers SO, Pazo V, Wright A, Bates DW. Preoperative hyponatremia and perioperative complications. Arch Intern Med 2012;172(19):1474-1481. [doi: 10.1001/archinternmed.2012.3992] [Medline: 22965221]
3. Allen SJ. Marker or mechanism? Dysnatraemia and outcomes in the perioperative period. Br J Anaesth 2016 Feb;116(2):155-157 [FREE Full text] [doi: 10.1093/bja/aev446] [Medline: 26787784]
4. Waikar SS, Mount DB, Curhan GC. Mortality after hospitalization with mild, moderate, and severe hyponatremia. Am J Med 2009;122(9):857-865 [FREE Full text] [doi: 10.1016/j.amjmed.2009.01.027] [Medline: 19699382]
5. Leung AA, McAlister FA, Finlayson SR, Bates DW. Preoperative hypernatremia predicts increased perioperative morbidity and mortality. Am J Med 2013;126(10):877-886. [doi: 10.1016/j.amjmed.2013.02.039] [Medline: 23910520]
6. Zhou XY, Chen HL, Ni SS. Hyponatremia and short-term prognosis of patients with acute pulmonary embolism: a meta-analysis. Int J Cardiol 2017;227:251-256. [doi: 10.1016/j.ijcard.2016.11.120] [Medline: 27839808]
7. Madsen CM, Jantzen C, Lauritzen JB, Abrahamsen B, Jorgensen HL. Hyponatremia and hypernatremia are associated with increased 30-day mortality in hip fracture patients. Osteoporos Int 2016;27(1):397-404. [doi: 10.1007/s00198-015-3423-4] [Medline: 26576542]
8. Ayus JC, Fuentes N, Go AS, Achinger SG, Moritz ML, Nigwekar SU, et al. Chronicity of uncorrected hyponatremia and clinical outcomes in older patients undergoing hip fracture repair. Front Med 2020;7:263 [FREE Full text] [doi: 10.3389/fmed.2020.00263] [Medline: 32695787]
9. Crestanello JA, Phillips G, Firstenberg MS, Sai-Sudhakar C, Sirak J, Higgins R, et al. Does preoperative hyponatremia potentiate the effects of left ventricular dysfunction on mortality after cardiac surgery? J Thorac Cardiovasc Surg 2013;145(6):1589-1594, 1594.e1-e2 [FREE Full text] [doi: 10.1016/j.jtcvs.2012.12.093] [Medline: 23566509]
10. Mohan S, Gu S, Parikh A, Radhakrishnan J. Prevalence of hyponatremia and association with mortality: results from NHANES. Am J Med 2013 Dec;126(12):1127-37.e1 [FREE Full text] [doi: 10.1016/j.amjmed.2013.07.021] [Medline: 24262726]
11. Corona G, Giuliani C, Verbalis JG, Forti G, Maggi M, Peri A. Hyponatremia improvement is associated with a reduced risk of mortality: evidence from a meta-analysis. PLoS One 2015;10(4):e0124105 [FREE Full text] [doi: 10.1371/journal.pone.0124105] [Medline: 25905459]
12. Chewcharat A, Thongprayoon C, Cheungpasitporn W, Mao MA, Thirunavukkarasu S, Kashani KB. Trajectories of serum sodium on in-hospital and 1-year survival among hospitalized patients. Clin J Am Soc Nephrol 2020;15(5):600-607 [FREE Full text] [doi: 10.2215/CJN.12281019] [Medline: 32213501]
13. Tzoulis P, Waung JA, Bagkeris E, Hussein Z, Biddanda A, Cousins J, et al. Dysnatremia is a predictor for morbidity and mortality in hospitalized patients with COVID-19. J Clin Endocrinol Metab 2021;106(6):1637-1648 [FREE Full text] [doi: 10.1210/clinem/dgab107] [Medline: 33624101]
14. Gu A, Chen FR, Chen AZ, Fassihi SC, Thakkar S, Unger AS, et al. Preoperative hyponatremia is an independent risk factor for postoperative complications in aseptic revision hip and knee arthroplasty. J Orthop 2020;20:224-227 [FREE Full text] [doi: 10.1016/j.jor.2020.01.028] [Medline: 32051674]
15. Pennington Z, Bomberger TT, Lubelski D, Benzel EC, Steinmetz MP, Mroz TE. Preoperative hyponatremia and perioperative complications in cervical spinal fusion. World Neurosurg 2020;141:e864-e872. [doi: 10.1016/j.wneu.2020.06.068] [Medline: 32553754]
16. Khan FW, Fatima B, Lahr BD, Greason KL, Schaff HV, Dearani JA, et al. Hyponatremia: an overlooked risk factor associated with adverse outcomes after cardiac surgery. Ann Thorac Surg 2021;112(1):91-98. [doi: 10.1016/j.athoracsur.2020.08.030] [Medline: 33080237]
17. Woo SH, Marhefka GD, Cowan SW, Ackermann L. Development and validation of a prediction model for stroke, cardiac, and mortality risk after non-cardiac surgery. J Am Heart Assoc 2021;10(4):e018013 [FREE Full text] [doi: 10.1161/JAHA.120.018013] [Medline: 33522252]
18. User guide for the 2018 ACS NSQIP: participant use data file (PUF). ACS NSQIP. 2018. URL: https://www.facs.org/-/media/files/quality-programs/nsqip/nsqip_puf_userguide_2018.ashx [accessed 2023-02-24]
19. Hawn MT, Graham LA, Richman JS, Itani KMF, Henderson WG, Maddox TM. Risk of major adverse cardiac events following noncardiac surgery in patients with coronary stents. JAMA 2013;310(14):1462-1472. [doi: 10.1001/jama.2013.278787] [Medline: 24101118]
20. Panos A, Mavridis D. TableOne: an online web application and R package for summarising and visualising data. Evid Based Ment Health 2020;23(3):127-130. [doi: 10.1136/ebmental-2020-300162]

XSL•FO
RenderX

21. Liu L, Nevo D, Nishihara R, Cao Y, Song M, Twombly TS, et al. Utility of inverse probability weighting in molecular pathological epidemiology. Eur J Epidemiol 2018;33(4):381-392 [FREE Full text] [doi: 10.1007/s10654-017-0346-8] [Medline: 29264788]
22. Pedersen EJ, Miller DL, Simpson GL, Ross N. Hierarchical generalized additive models in ecology: an introduction with mgcv. PeerJ 2019;7:e6876 [FREE Full text] [doi: 10.7717/peerj.6876] [Medline: 31179172]
23. Seaman SR, White IR. Review of inverse probability weighting for dealing with missing data. Stat Methods Med Res 2013;22(3):278-295. [doi: 10.1177/0962280210395740] [Medline: 21220355]
24. Lüdecke D, Bartel A, Schwemmer C, Powell C, Djalovski A, Titz J. Data visualization for statistics in social science. sjPlot. URL: https://cran.r-project.org/web/packages/sjPlot/index.html [accessed 2023-02-24]
25. Lüdecke D. ggeffects: tidy data frames of marginal effects from regression models. J Open Source Softw 2018;3(26):772 [FREE Full text] [doi: 10.21105/joss.00772]
26. Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York: Springer; 2009.
27. Kassambara A. ggpubr: 'ggplot2' based publication ready plots. R package version 0.1. ggpubr. 2018. URL: https://rpkgs.datanovia.com/ggpubr/ [accessed 2023-02-24]
28. Haneuse S, VanderWeele TJ, Arterburn D. Using the E-value to assess the potential effect of unmeasured confounding in observational studies. JAMA 2019;321(6):602-603. [doi: 10.1001/jama.2018.21554] [Medline: 30676631]
29. Mathur MB, Ding P, Riddell CA, VanderWeele TJ. Web site and R package for computing E-values. Epidemiology 2018;29(5):e45-e47 [FREE Full text] [doi: 10.1097/EDE.0000000000000864] [Medline: 29912013]
30. Localio AR, Stack CB, Griswold ME. Sensitivity analysis for unmeasured confounding: E-values for observational studies. Ann Intern Med 2017;167(4):285-286. [doi: 10.7326/M17-1485] [Medline: 28693037]
31. Vikas M, John K, Apruzzese P, Kendall MC, De Oliveira G. Utility of preoperative laboratory testing in ASA 1 and ASA 2 patients undergoing outpatient surgery in the United States. J Clin Anesth 2022;76:110580. [doi: 10.1016/j.jclinane.2021.110580] [Medline: 34794109]

Abbreviations

ACS: American College of Surgeons
ASA: American Society of Anesthesiologists
CHF: congestive heart failure
CPT: current procedural terminology
GAM: generalized additive model
NSQIP: National Surgical Quality Improvement Program
RR: relative risk

Edited by T Leung; submitted 06.04.22; peer-reviewed by V Guardabasso, C Weerth; comments to author 13.02.23; revised version received 19.02.23; accepted 21.02.23; published 16.03.23.

Please cite as:
Cole JH, Highland KB, Hughey SB, O'Shea BJ, Hauert T, Goldman AH, Balazs GC, Booth GJ
The Association Between Borderline Dysnatremia and Perioperative Morbidity and Mortality: Retrospective Cohort Study of the American College of Surgeons National Surgical Quality Improvement Program Database
JMIR Perioper Med 2023;6:e38462
URL: https://periop.jmir.org/2023/1/e38462
doi:10.2196/38462
PMID:36928105

©Jacob H Cole, Krista B Highland, Scott B Hughey, Brendan J O'Shea, Thomas Hauert, Ashton H Goldman, George C Balazs, Gregory J Booth. Originally published in JMIR Perioperative Medicine (http://periop.jmir.org), 16.03.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Perioperative Medicine, is properly cited. The complete bibliographic information, a link to the original publication on http://periop.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Toward Enhanced Clinical Decision Support for Patients Undergoing a Hip or Knee Replacement: Focus Group and Interview Study With Surgeons

Sabrina Grant[1*], BSc, MSc, PhD; Emma Tonkin[2*], MPhys, PhD; Ian Craddock[2], CEng, PhD; Ashley Blom[3], MD, PhD; Michael Holmes[2], PhD; Andrew Judge[1], PhD; Alessandro Masullo[2], PhD; Miquel Perello Nieto[2], MSc; Hao Song[2], PhD; Michael Whitehouse[1], MBChB, PhD; Peter Flach[4], PhD; Rachael Gooberman-Hill[1], PhD

[1]Musculoskeletal Research Unit, University of Bristol, Southmead Hospital, Bristol Medical School, Bristol, United Kingdom

[2]Digital Health, Faculty of Engineering, Bristol, United Kingdom

[3]Faculty of Medicine, Dentistry and Health, University of Sheffield, Sheffield, United Kingdom

[4]Intelligent Systems Laboratory, Department of Computer Science, University of Bristol, Bristol, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Emma Tonkin, MPhys, PhD
Digital Health
Faculty of Engineering
University of Bristol
Bristol, BS8 1TH
United Kingdom
Phone: 44 117 42 82343 ext 82343
Email: cselt@bristol.ac.uk

Abstract

Background: The current assessment of recovery after total hip or knee replacement is largely based on the measurement of health outcomes through self-report and clinical observations at follow-up appointments in clinical settings. Home activity-based monitoring may improve assessment of recovery by enabling the collection of more holistic information on a continuous basis.

Objective: This study aimed to introduce orthopedic surgeons to time-series analyses of patient activity data generated from a platform of sensors deployed in the homes of patients who have undergone primary total hip or knee replacement and understand the potential role of these data in postoperative clinical decision-making.

Methods: Orthopedic surgeons and registrars were recruited through a combination of convenience and snowball sampling. Inclusion criteria were a minimum required experience in total joint replacement surgery specific to the hip or knee or familiarity with postoperative recovery assessment. Exclusion criteria included a lack of specific experience in the field. Of the 9 approached participants, 6 (67%) orthopedic surgeons and 3 (33%) registrars took part in either 1 of 3 focus groups or 1 of 2 interviews. Data were collected using an action-based approach in which stimulus materials (mock data visualizations) provided imaginative and creative interactions with the data. The data were analyzed using a thematic analysis approach.

Results: Each data visualization was presented sequentially followed by a discussion of key illustrative commentary from participants, ending with a summary of key themes emerging across the focus group and interview data set.

Conclusions: The limitations of the evidence are as follows. The data presented are from 1 English hospital. However, all data reflect the views of surgeons following standard national approaches and training. Although convenience sampling was used, participants' background, skills, and experience were considered heterogeneous. Passively collected home monitoring data offered a real opportunity to more objectively characterize patients' recovery from surgery. However, orthopedic surgeons highlighted the considerable difficulty in navigating large amounts of complex data within short medical consultations with patients. Orthopedic surgeons thought that a proposed dashboard presenting information and decision support alerts would fit best with existing clinical workflows. From this, the following guidelines for system design were developed: minimize the risk of misinterpreting data, express a level of confidence in the data, support clinicians in developing relevant skills as time-series data are often unfamiliar, and consider the impact of patient engagement with data in the future.

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2018-021862

XSL•FO
RenderX

(JMIR Perioper Med 2023;6:e36172) doi:10.2196/36172

KEYWORDS

arthroplasty; knee replacement; hip replacement; orthopedic surgery; clinical decision-making; postoperative follow-up; home monitoring; wearables; video

Introduction

Background

Hip and knee replacements are major surgical procedures that aim to improve function and reduce pain related to joint diseases, particularly osteoarthritis. During hip or knee replacement, the affected joint is removed and replaced with an artificial joint. In 2019 in the United Kingdom, the National Joint Registry recorded 101,651 hip replacements and 108,713 knee replacements [1]; in the Unites States, >1 million total hip and knee replacement procedures are performed each year [2]. These surgical procedures are increasingly common, and numbers are projected to increase as a result of aging populations and increasing prevalence of obesity [3].

UK clinical guidelines for follow-up after hip and knee replacement surgery usually include face-to-face consultation, radiographs, and an assessment of health outcomes through telephone or web-based patient-reported outcome measures (PROMs) [4]. PROMs are designed to assess patients' own views of their health and outcomes without interpretation by clinicians or others [5]. Of these, generic measures such as the 12-item Short Form Survey [6] and EQ-5D [7] aim to assess all important dimensions of health-related quality of life [8]. The Oxford Hip Score (OHS) [9] and Oxford Knee Score (OKS) [10] are additional disease-specific PROMs used by orthopedic surgeons in the United Kingdom. As validated instruments, PROMs are valuable sources of information for clinicians and researchers. However, several practicalities must be considered when implementing PROMs: missing or incomplete data; potential burden for patients; and cost, time, and administrative labor-intensiveness [11-13]. A recent review found that PROMs were prone to several types of bias: bias because of collection mode; nonresponse bias; proxy or caregiver response bias; recall bias (eg, bias because of the quality of patient recollection of past states); language bias (eg, semantic ambiguity); timing bias, representing a limited number of snapshots; and fatigue bias [13,14]. The OKS and OHS in particular may also fail to stratify activity level across a younger, more active population as they are not designed for this purpose. Instead, other instruments such as the Knee Injury and Osteoarthritis Outcome Score may be used, which are specifically designed for young and physically active patients, capturing additional domains of sport and recreation function and knee-related quality of life such that it has greater responsiveness as an outcome measure [15]. Taken together, these issues mean that, although PROMs are extremely valuable sources of information for clinicians and researchers, particularly because of their standardized and validated status, it is worthwhile to consider other methods to assess outcomes after joint replacement [16,17] that could be used in parallel to PROMs to support decision-making.

In this study, we used a qualitative approach to explore how time-based data may be used by clinicians to supplement PROMs. Qualitative methods were used to explore inductively what matters to busy clinical staff and develop initial guidelines for a future system. Any system developed using these guidelines could be evaluated in future studies.

Measuring Activity in a Joint Replacement Population

An objective method of activity assessment—step counting—has been accurately used to monitor changes in gait and activity in musculoskeletal disorders and diseases affecting gait, including hip and knee arthritis [18,19]. The current objective method used to measure function is accelerometry via wearable sensors. These are inexpensive and easy to wear. However, the data currently derived from these sensors have some limitations, particularly when measuring complex activities and movements that are common in activities of daily living [17,20]. To capture the daily variation in a patient's functional abilities in their real living environment, it is necessary to move to automated measurement in the patient's home as well as toward analysis techniques that more directly reflect performance in activities of daily living. For example, cameras can be used to study the kinematics of the transition from sitting to standing [21].

The Sensor Platform for Healthcare in a Residential Environment (SPHERE) Interdisciplinary Research Collaboration has developed a technology comprising an integrated platform of low-power sensors that can measure information continuously about the home (eg, temperature, energy consumption, and humidity) as well as information about people in the home (eg, location, how active they are, and extent of movement) and their health-related behaviors [22]. Data capture has been demonstrated over months or years [23], and so the continuous time-series data collected by SPHERE offer a potentially useful source of data to supplement conventional methods such as PROMs.

This study considers real time-series data generated by SPHERE systems monitoring patient activity in the home before and after total hip or knee replacement. The types of data available from the SPHERE system in each home include metrics derived from Bluetooth-based indoor localization of the patient [18], continuous estimation of posture and ambulatory activities using a wrist-worn accelerometer [24], and silhouette data generated using a depth-sensing video camera [25]. Although the overall system was developed by SPHERE in the absence of equivalent commercial systems, the capabilities of such a system would readily be within the reach of several companies in the consumer *smart home* market. The costs of systems of this kind vary according to implementation decisions as different use cases may benefit from the deployment of different sensors. The patient burden is likely to be minimal once the system is successfully in place. In comparison with cross-sectional PROMs, the costs of continuous time-series data monitoring lie primarily in maintenance following initial installation, and hence, time-series approaches may be more practical for

XSL•FO
RenderX

longer-term observation of patients' symptoms. Therefore, the findings of this study are a good guide to the strengths, weaknesses, and potential clinical utility of a near product.

Objectives

The main objectives of this study were to (1) introduce surgeons to continuous home data by visualizing time-series sensor data, (2) understand how these data could assist in postoperative clinical decision-making, and (3) identify design recommendations arising from clinician feedback.

The study departs from previous literature on the use of data in clinical decision support, which is largely focused on data from clinical environments such as intensive care [26,27]. To date, studies that have presented data from community settings to surgeons have focused principally on manually clinician-reported data [28] and laboratory outcomes [29], such as those commonly stored in electronic health records, or patient self-reported data [30] such as PROMs [31]. Where sensor data are sampled, this is often at a relatively low sample rate (eg, a daily measurement or 12 measurements per day) or over a relatively short period, from a few minutes [32] to a week or month [33]. Herein, we consider the challenges of how busy surgeons would make sense of thousands of data points a day over periods of up to 3 months—as would be easily within the capability and requirements of a home-based sensor system [34,35] monitoring recovery from major surgery [36].

Methods

Participants and Recruitment

From October 2018 to May 2019, orthopedic surgeons at a hospital in South West England, United Kingdom, were invited to take part in a focus group study. Participant demographics were collected (sex and level of experience performing hip or knee replacement surgical procedures). Identification of potential participants was conducted using convenience and snowball sampling. During this process, surgeons known to the study team were asked to identify other potential participants. Participants were initially screened against the inclusion criteria (a minimum of 2 years of experience performing total hip and knee replacement procedures). The exclusion criteria were a lack of experience in joint replacement specific to the hip or knee. Potential participants were emailed invitations, and those who agreed to consider taking part were invited to attend focus groups. Individual interviews were offered if the focus group timings were not suitable. Those who were contacted were also asked to nominate other potential participants—a snowball sampling approach.

A total of 9 participants (surgeons who were either working as consultants or registrars [residents]) took part. Several potential participants declined because of their clinical workloads and time constraints. At the start of each focus group, the study was discussed with the potential participants, who were invited to ask any questions about the study. Before the focus group started, they provided their written informed consent to participate, including to the publication of anonymous quotations. Focus groups were held at a clinical research center on the same site as the hospital to make it as straightforward as possible for busy surgeons to attend [37]. Face-to-face interviews at the surgeons' places of work were offered where attendance to the focus group was impractical because of time or distance. In total, most of the surgeons (7/9, 78%) attended 3 focus groups, and 22% (2/9) attended one-to-one interviews. The sample size was considered adequate as enough information was collected to clearly demonstrate concepts or ideas related to the topic addressed and with sufficient repetition of those concepts [38].

Ethics Approval

Ethics approval was provided by Southwest – Central Bristol National Health Service (NHS) Research Ethics Committee (17/SW/0121) on June 22, 2017.

Topic Guide and Procedure

A structured topic guide (Multimedia Appendix 1) was developed by the research team, which comprised an interdisciplinary group of health researchers and psychologists (SG and RGH), orthopedic surgeons (AB and MW), data scientists (IC, HS, ET, MP, AM, MH, and PF), and a translational statistician (AJ).

In part 1, a scenario (Textbox 1) was used as a tool to explore the current clinical systems in orthopedic care.

In part 2, a series of visualizations (Figures 1-9) were presented based on real participant data from the SPHERE 100 Homes study [23] (in which the system was deployed in homes of the general public) and the Hip and Knee Study of a Sensor Platform for Healthcare in a Residential Environment [39] (in which the same system was deployed in the homes of orthopedic patients).

Textbox 1. Scenario 1—Joyce (aged 63 years).

- Joyce is a 63-year-old lady who lives in a large three storey house with her daughter and daughter's fiancé. Joyce is a self-employed therapist and runs her practice from her home. She has a second part-time role at the local University as an administrator.
- Joyce previously had trouble walking distances. Because of a limp she uses a walking aid at times and reports significant hip pain.
- Joyce has recently had her left hip replaced.

Many metrics can be generated using home sensor data. For the purposes of this study, a series of target metrics were generated based on the literature on hip and knee studies. Metrics referenced in manually administered survey instruments such as the OHS and OKS [10] and the Pittsburgh Sleep Quality Index [40] were considered useful targets. Once this step was complete, a series of sample visualizations was generated using these metrics. These visualizations were first proposed and improved over multiple discussions and careful analysis of real patients by members of the Hip and Knee Study of a Sensor

XSL•FO
RenderX

Platform for Healthcare in a Residential Environment—mainly data scientists and health researchers. After several iterations, the resulting visualizations were used as examples to provide during the focus groups with clinicians. The detailed rationale behind the development of these figures has been published separately [24]. The use of realistic (eg, noisy and incomplete) prototype data from real homes and real patients was considered desirable throughout this study to ensure that the feedback was related to the characteristics of achievable systems that could plausibly be developed for clinical use in the future.

Scenario-Based Exploration

An action research approach was used in which participants were seen as able to identify value in context when encouraged to take initiative and identify possibilities for improvements [41].

Participating orthopedic surgeons were presented with a fictitious but realistic orthopedic patient scenario (Textbox 1); the narrative nature of the scenario approach is known to be a useful tool in the design process [41]. Surgeons were asked to focus on the 6- to 8-week postoperative consultation for this hypothetical patient. This creates a familiar and meaningful context [42] in which they are well placed to imagine whether new forms of data would assist them in carrying out their professional responsibilities.

To maximize discussion and allow participants to write thoughts and views independently, participants were provided with printouts of presented visualizations for use within idea generation sessions.

Focus Groups and Interviews

A total of 3 focus groups (with 2-3 surgeons in each group) and 2 interviews were conducted to explore the data visualizations. Each focus group was facilitated by 2 researchers and lasted approximately 1 to 1.5 hours. The interviews lasted approximately 45 to 60 minutes. The focus groups and interviews were digitally audio recorded and transcribed verbatim.

Part 1 of each data collection phase discussed the scenario (Textbox 1) exploring the assessment of recovery for patients after surgery.

Part 2, led by data analysts (MH and MPN), was a structured exploration in which the surgeons were presented with a selection of visualizations based on data generated from the homes of 2 orthopedic patients who had been recovering from total hip replacement. Participants were asked to consider the use of the visualizations as a way of assessing patient outcome and recovery after surgery. Participants were provided with paper copies of each visualization for any further thoughts or comments that were not captured by discussion—these were reviewed during the coding process.

Data Analysis

Qualitative data from the focus groups and interviews were analyzed using an inductive thematic approach [43]. The initial labeling generated a list (a “frame”) that was then systematically applied to the data and refined as the analysis progressed [44]. Members of the research team from clinical and nonclinical disciplines were allocated 10% of these transcripts to label independently. After collaborative discussions, further labels were identified, defined, and grouped into themes. This process of investigator triangulation increases internal validity [44]. Excerpts of data were placed on charts according to themes. All data were managed using NVivo software (version 12.0; QSR International).

This qualitative study focuses on the views expressed by the surgeons. Quantitative data were presented to surgeons to elicit those views, but the quantitative data themselves were not the subject of this study. A brief description of the method used to generate the quantitative data and visualizations is provided in the Results section, and the interested reader is referred to the study by Holmes et al [24] for further details.

Once participant feedback was evaluated and themes were identified, feedback was presented to the interdisciplinary research team consisting of researchers, surgeons, machine learners, and interface engineers. This step was intended to facilitate the integration of these findings into future iterations of the sensor and data analysis platform. This step resulted in the development of a series of guidelines that integrate findings from the participants with insights from the interdisciplinary team. This asynchronous codevelopment approach offers an opportunity for participant surgeon feedback and guidance to be made available for future engineering and design processes through the provision of guidelines.

Results

Participants and Recruitment

A total of 9 surgeons agreed to participate. Of these 9 surgeons, 6 (67%) were consultant orthopedic surgeons and 3 (33%) were orthopedic registrars (residents)—all the participants saw patients and conducted hip or knee replacement surgery as part of their usual workload, with experience ranging from 2 to 25 years. In total, 22% (2/9) of the participants were female, and 78% (7/9) were male.

Scenario-Based Exploration

The participants were led in a scenario-based exploration via a fictitious patient scenario, as described in Textbox 1. This generated a stimulated and creative discussion among participants, the outcomes of which are presented in the following section. During data collection and analysis, it became clear that there was a reasonable degree of agreement and repetition in the findings, and sufficient information was deemed to have been obtained in relation to the subject area. In light of this, recruitment and data collection were stopped once 9 surgeons had taken part [38].

Outcomes From Focus Groups and Interviews

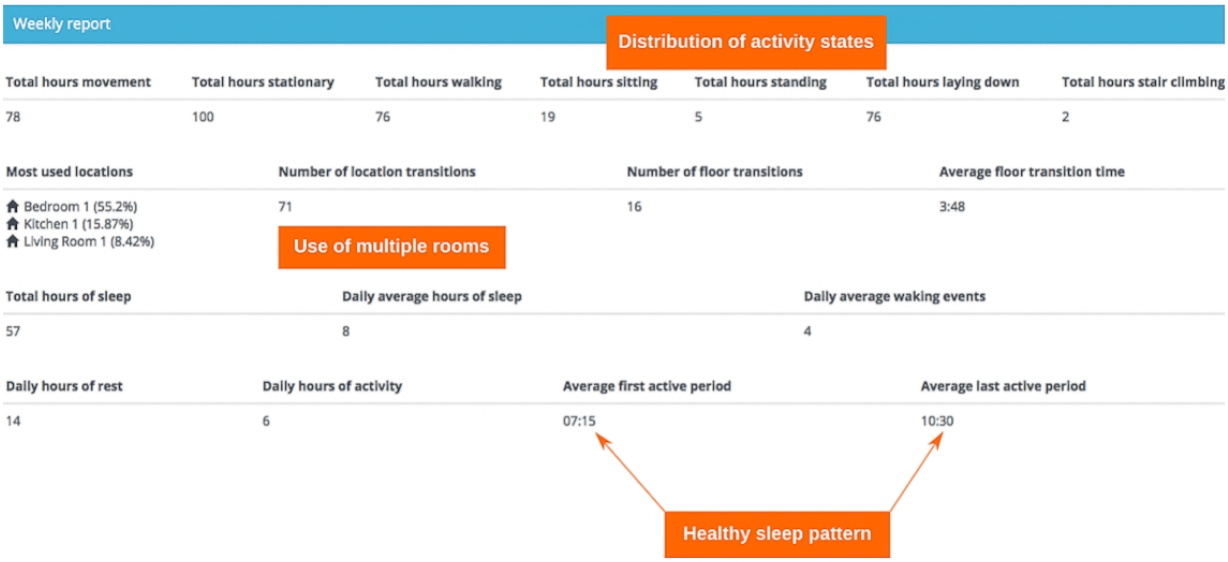
Individual commentary from participants is presented regarding each of the visualizations (Figures 1-9), followed by a series of broader themes from the focus groups and interviews.

Visualization 1 (Figure 1) presented a series of summary statistics calculated using patient indoor location and accelerometer data. These included room-level occupancy data,

XSL•FO
RenderX

transitions between rooms, activity predictions generated via machine learning, and actigraphy analysis. The sample data given here were intended to be representative, not exhaustive, and it is possible to identify many other summary statistics that could be relevant.

Figure 1. Visualization 1—a screenshot of a dashboard displaying summary statistics, including activity, location, room transitions, floor transitions, and sleep routine.



Participants stated that, although the tabulated data looked very informative, rapid extraction of relevant information was not straightforward given the constraints of a busy 10-minute clinic consultation. Population norms or other references would be required to assess any change or improvement. Visuals would be improved if areas were highlighted to provide a focus for surgeons during a consultation:

> *So I don't know where the mobility is on total hours walking, 76, I don't know what that means, is 76 a lot, is [it] not a lot? Obviously if you had pre-op and post-op data then that's great because you could get the data just to show you which are better, worse, whatever, but that I think I wouldn't look at because it would be too hard to navigate.* [#0046]

Participants reflected on how current methods of assessing patient health outcomes using the OHS or OKS lacked some of the valuable temporal information contained within the SPHERE time-series data. It was suggested that this insight into function over time could help them better understand the recovery process.

Most participants (7/9, 78%) suggested that the existing routine (face-to-face) clinic follow-up appointment presented a similarly rich opportunity for assessing recovery through movement. See the Themes Arising From Data Analysis section for discussion of this point:

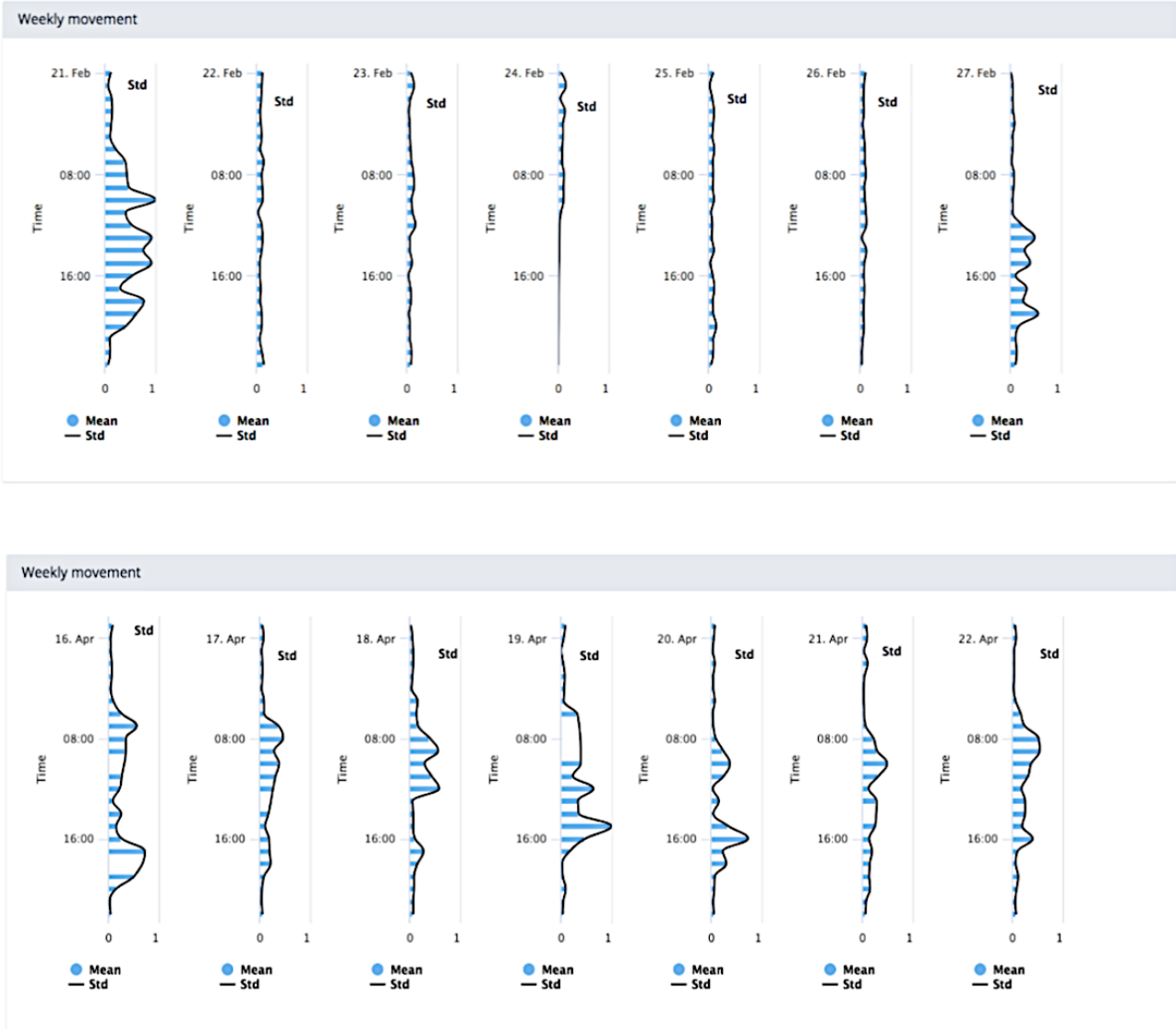
> *So I think actually the bits that I think are important on here all come back to Oxford Hip Score. So people getting up and down stairs, people on the move, people sleeping you know, these are all things that are kind of covered in one way and that, but this [the sensor data] gives you more detail, it's not just "Yes, good, very good" or whatever. So for me I think, although like the moving one [visualization 7] is quite cool...I don't see how me looking at that no matter how many hours I had to look at the patient, that when they come in and they stand up and sit down I've made my judgment whether they've got a problem or not.* [#0050]

> *I suppose I'm expecting a patient to be compared against other patients that I've seen before. Usually, when I see them, the first thing I do is watch them walk. I watch them walk into the consultation room and check whether or not they're using walking aids. I then take a good history about how their recovery is going and whether that's meeting their expectations as well as mine.* [#0051]

Visualizations 2 and 3 (Figures 2 and 3) illustrated the recovery of 2 patients over weeks 1 and 6 after surgery. The top and bottom figures correspond to weeks 1 and 6, respectively. Patient 1 was an example of a "good recovery," and patient 2 was an example of a "poor recovery." "Good" and "poor" recovery were determined descriptively from analysis of PROM data on function and sleep and qualitatively from interviewing the patients, which were data collected as part of the wider program of study [45,46].

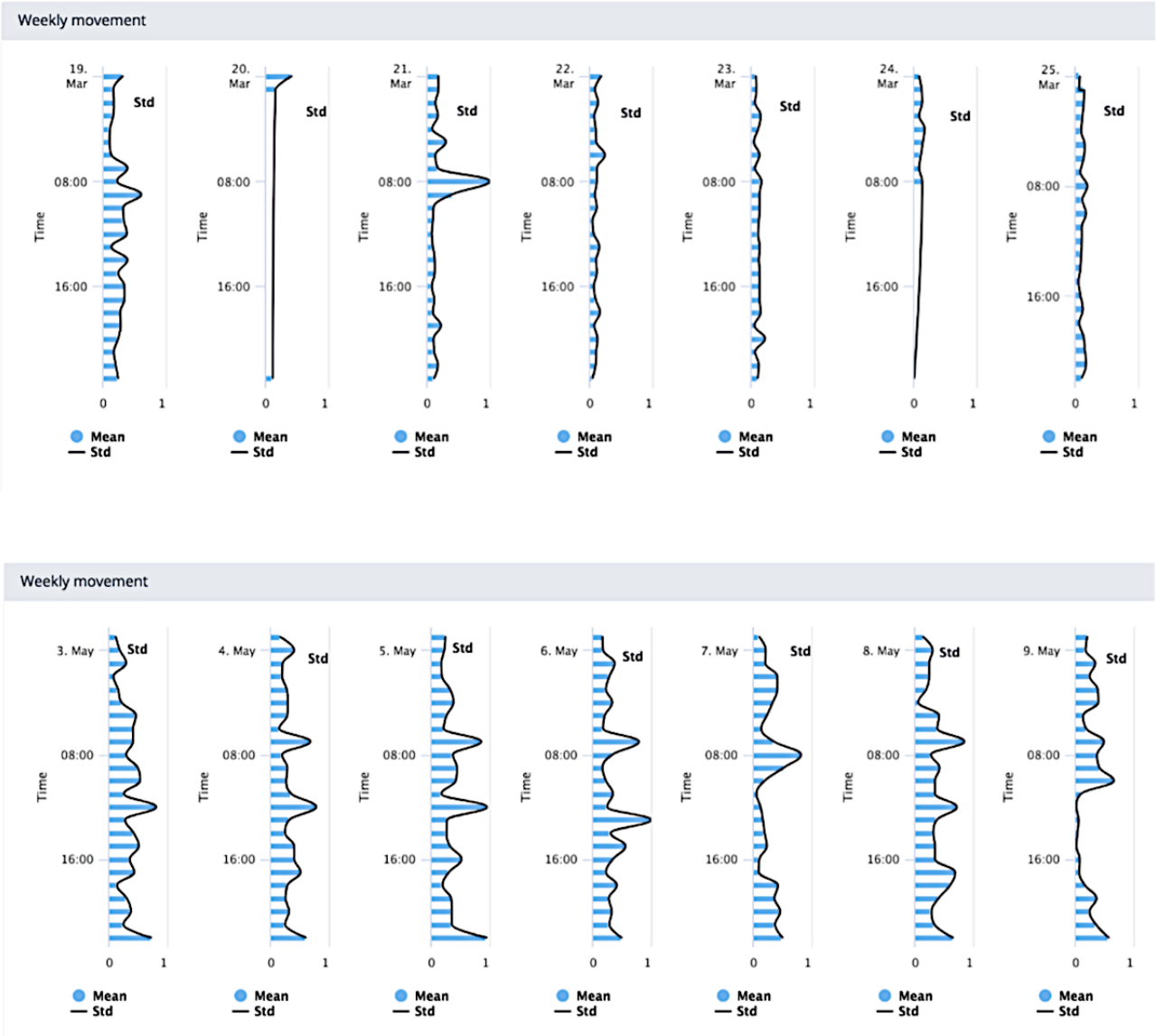
XSL•FO
RenderX

Figure 2. Visualization 2—activity levels of a patient recovering well after surgery.



XSL•FO
RenderX

Figure 3. Visualization 3—activity levels of a patient recovering poorly after surgery.



The visualizations made use of accelerometer magnitude data, preprocessed to establish the SD, which is an approximate measure of physical activity [47]. These data were then visualized as a series of horizontally stacked axes, each representing a day of the week. Data for 1 week were available in each chart. Stacked charts are a commonly used approach to time-series visualization, with known limitations; notably, line length is visually easier to compare than position [48], and hence, comparison of multiple series is challenging using this type of chart.

Some general suggestions were offered to improve visualization 2. Displaying the mean value of that week rather than daily values was suggested as easier to use:

> *Because if you put all that together you would get a trace that resembles that [visualization 2/3] and you'd be able to say straightaway that at six weeks, they're doing great and then at one [week], they're not.* [#0046]

However, it was noted that striking the right balance of information using the mean would be a challenge:

> *The problem is, if you average everything out, then you lose the detail don't you...but if you present all of the detail, then it becomes impossible.* [#0046]

Adding a reference value was suggested by some as useful within the weekly charts to demonstrate a "typically" poor recovery and where the patient sits in relation to that. Visually representing this recovery process for patients was again considered a better outlet for a conversation regarding the expectations of surgery:

> *I guess in the second slide [visualization 2-3] if you were able to put some points to say, this is low activity at time of sleep which is what you're expecting...* [#0051]

However, there was a level of disagreement between some surgeons about using population-based comparators with patients, that is, comparing a patient's data with population norms or averages:

> *No I think it would depersonalize it in my opinion...Your [the patient's own] starting point...may be very, very much lower than another patient. So I*

XSL•FO
RenderX

> *would just work on an individual improvement.* [#0050]

The assessment of sleep patterns across a 12-week period is beyond the conventional self-reported assessment of sleep. Within these focus groups, participants viewed long-term changes in sleep as a new approach to understanding recovery. However, as it was an unfamiliar metric, there were varying opinions about how useful this could be for making clinical decisions. On the one hand, it may be helpful for more tailored advice:

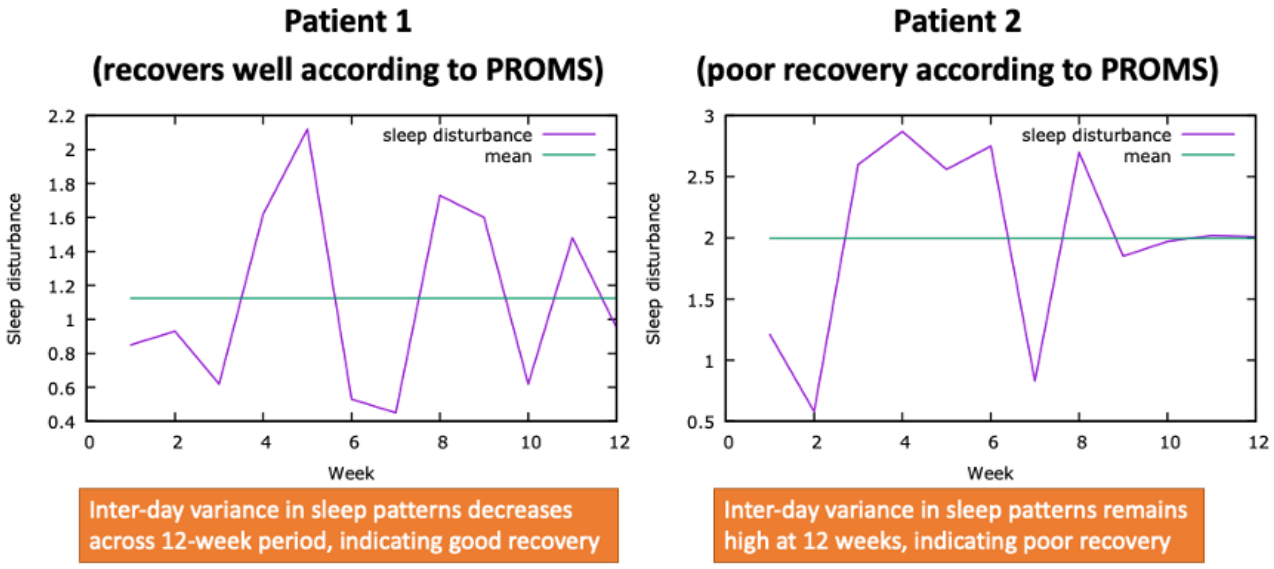
> *Postop [after surgery] we tell them, they've got to sleep on their back and most of them have got a bad back, and they hate it, so that's why they're doing this but this would be so interesting if in time we changed to give them advice to sleep any which way they like...you might notice a real difference.* [#0049]

In contrast, some felt that the minutiae of sleep were affected by several different factors following surgery and, therefore, could not be assessed or considered alone:

> *There are so many other factors that are going to affect sleep other than the joint replacement, especially in this particular demographic...There are potentially too many confounding factors in there for us to be able to use it [visualization 4] usefully.* [#0051]

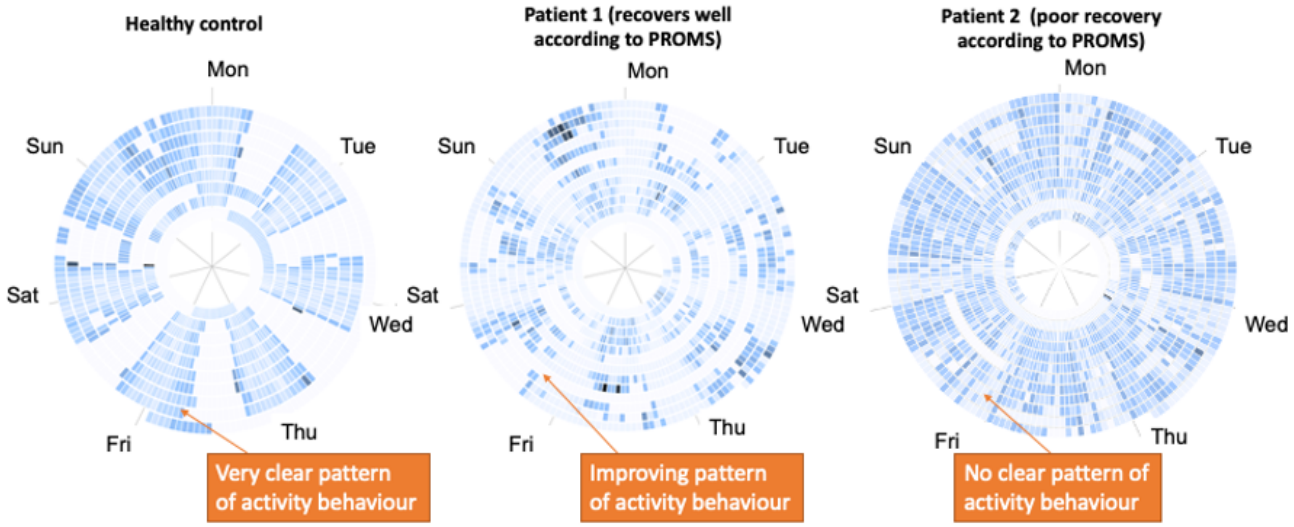
Visualization 4 (Figure 4) graphically represents intraday variance of sleep patterns drawn from raw actigraphy data.

Figure 4. Visualization 4—sleep trend data. PROM: patient-reported outcome measure.



Visualization 5 (Figure 5) presented a spiral representation of patient physical activity (derived from accelerometer data). Each complete ring represented a week of data—the spiral map was chosen in response to the observation by Weber et al [49] that spiral charts permit visualization of lengthy time series and that the circular representation is appropriate for time series with high periodicity (in this case, weekly periodicity). Each chart represented approximately 2 to 3 months of data.

Figure 5. Visualization 5—long-term interpatient comparison of movement. PROM: patient-reported outcome measure.



XSL•FO
RenderX

The color scheme applied identified absence of data as pure white and represented increasingly strenuous activity using darkening shades of blue. Hence, regular and restful sleep patterns were reflected using light blue "striping" through each period of least activity, with the most active periods represented as dark areas. Periods outside the home resulted in white striping (absent data) on the chart as the project did not collect data outside the home. Known limitations of the spiral format include difficulty reading "older" data at the center of the spiral because of the small area of the central cells. Facilitation involved some narrative running alongside the presentation of visualizations 5 and 6, which explained the unusual visualization to participants.

Across the focus groups, opinions varied on the use of these representations.

In contrast to tabulated data illustrated in visualizations 2 and 3 (Figures 2-3), participants felt that this visualization provided a layer of depth to understanding activity beyond a conventional table:

> *I quite like...[visualization 5] because I think that at least gives you a bit more depth to the data rather than just a bog-standard table form...visually the patient can understand it with a similar explanation.* [#0052]
>
> *...you want it to be as simple as possible, the patients want to see [physical] models, they want to see x-rays, they want to see a simple form of data that shows that they've done better, or they are improving.* [#0050]

Although interesting to some surgeons, most participants (6/9, 67%) thought that simple line graphs would be more user-friendly.

Furthermore, comparing data from week to week seemed to be favored as a tool for discussion with patients, primarily for illustrating any improvements to them:

> *...again we want to have less explanation to the patient as possible, so something visual that they can see, this was my activity level, a percentage even, pre [surgery], this is what is was post [surgery]...the patients just want to know has it made any difference, has it improved from their pre-operative state?* [#0050]

Presenting data in this way stimulated new ways of thinking about activity for the participants—specifically, looking at variation in activity levels over time rather than absolute magnitudes:

> *I think it's brilliant what it's capturing in the house!* [#0049]

Participants indicated that the variability visible in the charts was of interest but that it was difficult to interpret:

> *This is an unusual way to display data.* [#0051]

Making incorrect inferences from the data within a consultation was a concern for the participants, and therefore, guidance or training would be needed for surgeons to use these unfamiliar visualizations:

> *During the investigation with the patient I would not use the wheels [visualization 5/6]...because it would take twenty minutes to explain to them and half of them still wouldn't understand. It's quite a difficult concept.* [#0049]

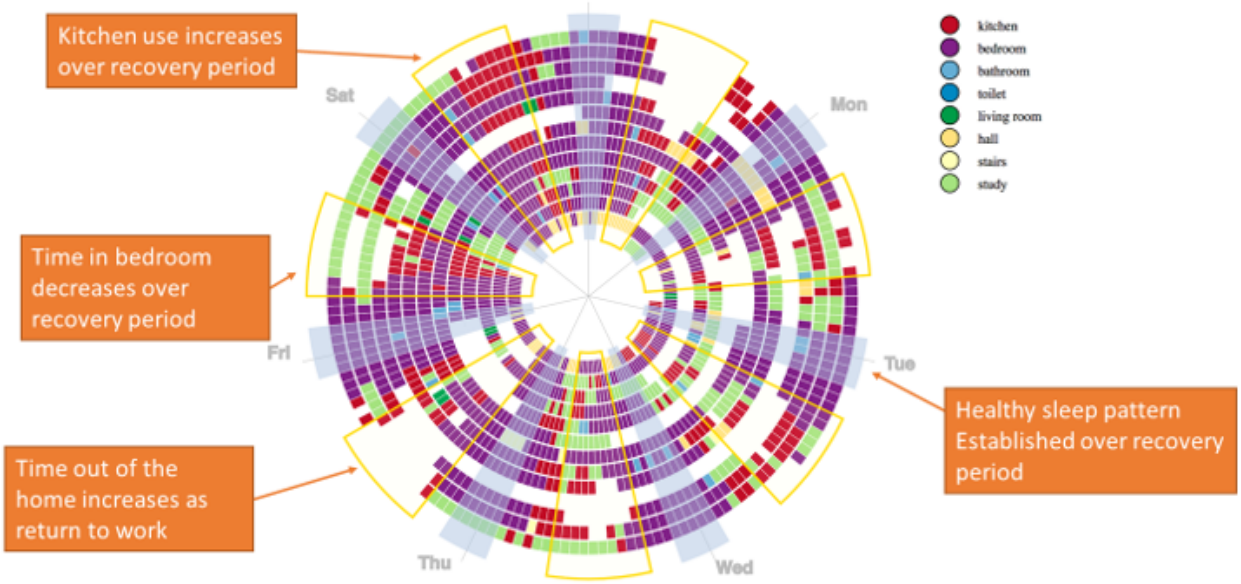
Nevertheless, there was broad agreement that such visualizations would be useful for surgeons to use ahead of the consultation but not to explain to patients:

> *I can see that that [circular plot 1 in visualization 5] is regular and yeah, great, and I can see that [circular plot 2 in visualization 5] is somewhere in between...but I wouldn't be able to interpret what the hell that means.* [#0046]
>
> *But looking at that and understanding it, I like it but having it explained, having that as a visual reference with the patient in clinic, it would take too long.* [#0049]

Visualization 6 (Figure 6) presented a spiral chart [49] designed on the same principles as visualization 5, representing a summary of room occupancy information drawn from 14 weeks of a patient's recovery. Estimates of the average least active times (ie, "L5" in actigraphy terms) and most active times in the participant's day were drawn from actigraphy data [50] and applied to the chart as an overlay to guide the eye to sections that were expected to have similar characteristics. Empty cells indicated that the participant was not present in the home at that time.

XSL•FO
RenderX

Figure 6. Visualization 6—indoor location data.



Visualization 6 captures trends in participant behavior within the home. Although clinicians thought that it was useful to describe these trends, they also considered that the data were too complex:

> *The trends are useful, but although you can see the trends on there, they don't jump off the page.* [#0052]

However, identifying anomalies or information that conflicted with any predicted outcomes or expectations was useful:

> *If you saw that and it shows that they're not spending any time in the kitchen because they're immobile because they can't walk, then that's useful.* [#0051]

However, the length of time to arrive at these conclusions was an issue in a short consultation per patient:

> *It's really interesting but it takes a long time specifically to, I guess, actually put meaning to it...it needs to highlight people that are struggling or not getting on, rather than presenting really intricate data about what they're doing which is interesting but...I couldn't sit down and look at that with every post-op patient.* [#0046]

Sharing this information with the patient was further highlighted as a challenge if patients interpreted it incorrectly:

> *That is just an absolute bombardment of colour and data to a patient. It would take you ten minutes to explain activity levels, trends, patterns.* [#0051]

Accurate assessment of activity using concentric circles was also an issue for some clinicians:

> *I have a little bit of a problem with it being displayed as a circle because the radius of the circle increases. The surface area of each block increases as you go out. I think it looks like that amount of time is less because you're looking at something closer in.* [Participant 1 in 0051]

> *It feels a little bit as though we're exaggerating the good bits on the outside.* [Participant 2 in 0051]

Capturing change over time was viewed as an essential component of the data, and this was not met by this visualization:

> *I mean, largely, our job is looking at change over time. I don't think you can interpret change over time very easily on that, I would say (Participant 1 in 0051)...I think we want something that quickly conveys the information that is most important to what your clinical decision making will be at that point.* [Participant 2 in 0051]

The sit-to-stand movement is used to assess patients in clinics and in research. It can be evaluated using a variety of metrics, including the speed of the motion [51]. Figure 7 is a screenshot of a video showing multiple sit-to-stand transitions collected over the progression of the patient's recovery (this was an early result from the project, and the video has since been much improved). The data are ordered from left to right, with more recent data to the right.

SPHERE does not store videos captured in people's homes, and hence, the visualizations presented are "silhouettes." The use of silhouettes was designed to ensure privacy and acceptability (including acceptability to nonpatient household members and visitors to the home), and some similar processing is a likely feature of any commercial product developed for a similar purpose.

Although the moving images were pleasing, participants least preferred this visualization. Most surgeons expressed that this transition is, in most cases, informally assessed by them as a matter of routine as the patient comes into the first follow-up consultation:

> *If they walk into your room and they can't stand up or sit down you know the answer. You've actually clocked it before they've even got into your room because you've watched them get up in the waiting room and walk towards you and whether they've got*

XSL•FO
RenderX

> *sticks and things. So yeah and you don't really care about the trend, in that one it's the absolute.* [#0052]

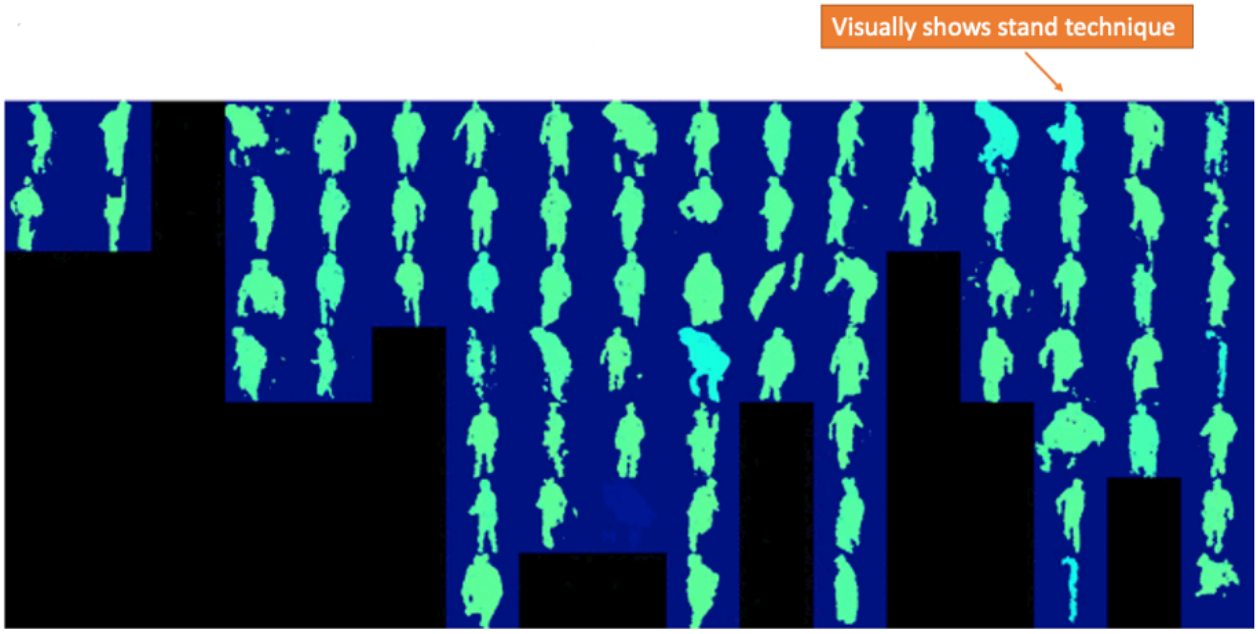
Participants recognized that visualization 7 (Figure 7) presented an early prototype of movement data display and not a final outcome:

> *I don't think there's enough resolution there to understand the sit to stand process...we're primarily concerned about are they flexing too deeply in the early phases, are they rotating and you can't really see rotation there.* [#0049]

Sharing this information with patients was also problematic:

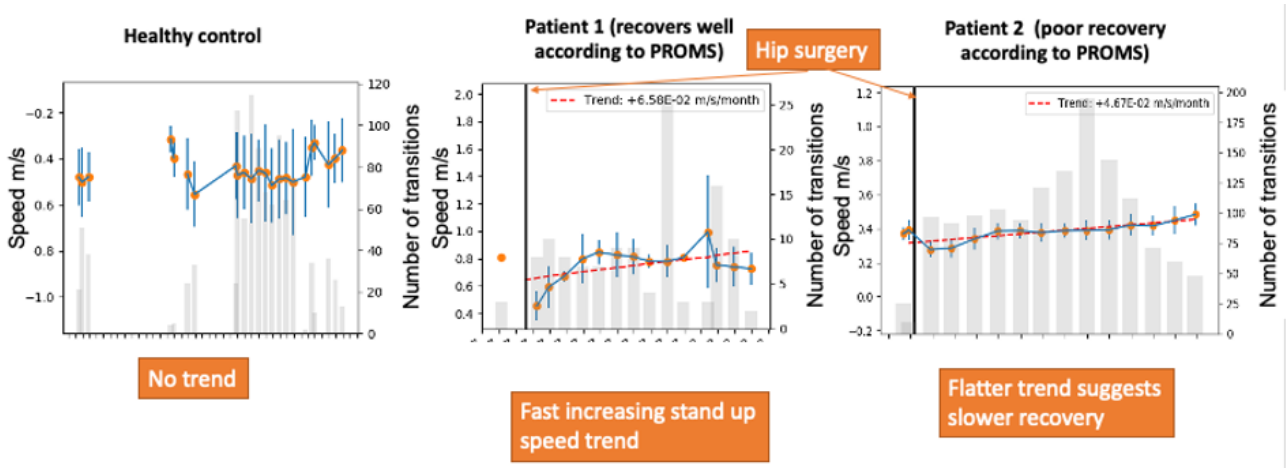
> *It's too many different images flashing at the same time for a patient to make heads or tails of it.* [#0050]

Figure 7. Visualization 7—movement data in long-term sit-to-stand transitions over 16 weeks.



Visualization 8 (Figure 8) presented the progression of a quantitative sit-to-stand metric—average speed—over several weeks of recovery time. This is automatically extracted from the data presented in visualization 7 and, hence, constitutes a simplified view of those data.

Figure 8. Visualization 8—trends in sit-to-stand speed. PROM: patient-reported outcome measure.



The speed of transitions over time was viewed as helpful, although an extension of this line of inquiry would be to look at daily habits changing over time:

> *But not necessarily the sit to stand times so much as actually I was thinking "Oh it'd be interesting to know whether they get down the corridor to the kitchen quicker" because most people having a hip replacement will have a seat they sit in during the day and they will go and make cups of tea and it would be interesting to know are they faster at getting to their kettle over six weeks.* [#0052]

The speed of the sit-to-stand transition over time is an existing objective test that is sometimes used in clinics. Therefore, this measurement is not in itself an advance in the state of the art in assessing recovery. Rather, the innovation in this case is its use as an in-home metric collected daily. Surgeons proposed that a

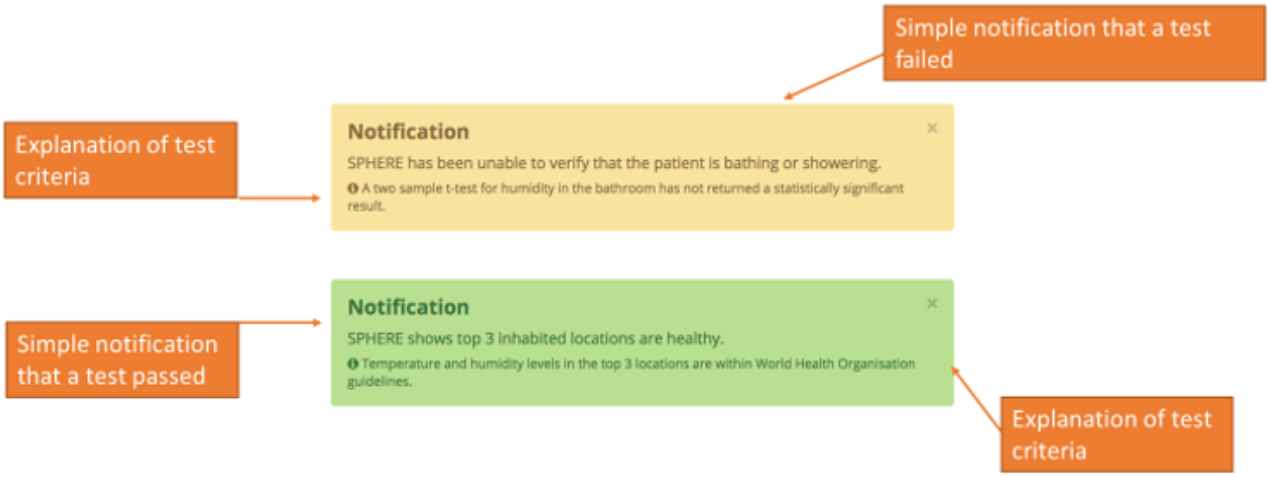
XSL•FO
RenderX

broader range of metrics could be used to account for the observed range of patient behaviors:

> *Your data over time from sitting to rising [visualization 8] is useful. There are a couple of other tests so that's very useful, [an] easy graph for patients to understand. But that's just one specific activity that you're looking at.* [#0050]
>
> *Yeah, but it's a pity it doesn't measure how far they walk, and is it possible to capture activity data outside the home? Because we do get patients who will [unnecessarily] restrict themselves, particularly older patients, doing their exercises in the house and you'll get patients at week one or week two, [who] are going round the block.* [#0049]

Visualization 9 (Figure 9) presents a screenshot of a prototype decision support tool. Sample notifications are presented that make use of available environmental and participant localization information (such as that displayed in Figure 6) to generate responses to 2 example tests: patient bathing or showering (top) and the suitability of the environmental conditions within the home in comparison with standard guidelines (bottom). In practice, it is likely that many unitary tests of this nature exist. Hence, to avoid a "busy" interface, the results would be filtered in accordance with clinical decision support recommendations in a real-world context of use [52].

Figure 9. Visualization 9—decision support data notifications and alerts.



This final visualization stimulated lively discussions in the focus groups and interviews. The presentation of a notification dashboard aligned with several of the clinicians' previous expectations of how the data might look. Participants consistently mentioned that clinical tools should capture key pieces of information that can be interpreted rapidly and accurately by both the clinician and patient:

> *I think that's what most surgeons would use, and they'd have ten minutes in the clinic and they go.* [#0049]
>
> *I really like the dashboard, "your patient can't sleep," "your patient can't do the stairs," "the humidity suggests they haven't had a shower for two weeks," you know, that kind of data is really helpful—"they can't cook."* [#0046]
>
> *This is a very good thing, because you don't want the surgeons interpreting their own way because they might interpret very differently.* [#0049]

Therefore, the presentation of this visualization best met the expectations for a tool that could be used in current clinical workflows:

> *When we go back to that one [visualization 9] that's quite useful is someone automatically telling us, exactly as you said there, a flag saying "the patient isn't going out as much as they used to" or "the patient is going out more than they used to, the patient doesn't appear to be sleeping as much as they used to" or "sleep patterns are still irregular at 6 weeks." I think those notifications would be good, but I could imagine there being a fairly hefty list of them.* [#0052]
>
> *You know, much as we get ten minutes per patient, you will get the occasional patient who'll take forty minutes to sort out with a ten-minute slot. And then you're playing catch up for the next five patients, and that's the point where you switch to the notification screen, right, is there anything standing out that I need to know about.* [#0049]

Textual summaries such as these notifications alerting the surgeon to relevant features within the data were perceived as aligning well with a 6-week follow-up consultation appointment routinely offered to patients during which the surgeon could include this information in their existing review procedures. Central to this appointment was the opportunity for surgeons to identify any early warnings of surgical complications from the patient's perspective:

> *I think it is useful to have these notifications definitely, the point of the six week follow up is to identify patients who've got a problem that we need to do something about and those are principally early infection and dislocation or an early fracture. That's the point of that six week appointment and so that's what we need these tools to tell us. Or [alternatively]*

XSL•FO
RenderX

> *the other point of it would be do we need to offer more social support or physiotherapy support in this patient's home to prevent them falling over or help them?* [#0051]

> *At six weeks, its actually quite useful because if they're not actually showering, they might think that's [be]cause they can't get the wound wet, patients do have concerns about that.* [#0049]

Participants broadly liked the idea of a top-level decision-making process integrated within a dashboard, primarily as this removed the need to interpret data for every patient and enabled integration more easily with current clinical management systems:

> *So you mentioned dashboard, so if I was on BlueSpier [a clinical management system]...if I went in to the patient...if it came up as a red alert, then I'd have a look at it and then go into the data a little bit closer and speak to them about it.* [#0046]

Some participants (3/9, 33%) felt that similar reports could usefully be distributed in paper form (or presumably by email) to the patient, suggesting that patients may engage with this information outside appointments:

> *If there was an available one [printout] absolutely, patients would be able to take it home and have a look at it.* [#0050]

Of the visualizations presented, visualizations 2 to 8 were considered better suited as research tools, whereas the suggested dashboard and notification system presented in visualization 9 (Figure 9) was more appealing for clinical use:

> *As a clinical tool, I think the notifications are very helpful. I think what would be useful is if you actually provided it to us and gave it to a few surgeons and test it.* [#0049]

Themes Arising From Data Analysis

Overview

Several general themes were identified common to all visualizations. These were generated by thematic analysis as described in the *Data Analysis* subsection of the *Methods* section. Four themes common across focus groups and interviews were (1) home data represent a more objective measurement of activity, (2) home data provide a stimulus for discussion in a consultation with a patient, (3) there is interest in the use of home data for clinical research purposes, and (4) there is a need to meet clinicians' requirements in the development of visualizations.

Data From the Home Can Give a More Objective Measurement of Activity

Assessment of a patient following surgery is mostly done via a face-to-face clinical appointment approximately 6 to 8 weeks after the operation. In this appointment, through questioning, surgeons routinely assess how patients are getting on at home with their activities of daily living and general independence. Participants widely felt that the SPHERE home data presented an advancement from this current practice:

> *What would be quite good with this is that you get an objective measure so, you know, can you cook? "Oh yeah, much better." But you're not cooking, so it's not better, maybe physically you can't cook but the times you have managed to get into the kitchen and it didn't hurt so you remember it as being better but you've only cooked one meal in the week.* [#0049]

> *So assessing your patient, you take them on their word really as to how they're doing. So you ask about how they're getting on at home, activities, are they still independent, is someone else doing the shopping, can they manage stairs, have they moved, so it's all those sorts of things which we take on their word. So I suppose this would give you objective information as to whether that's true, not saying that they're not telling the truth but it would just give you another side of things.* [#0050]

> *The only thing that this offers that we struggle with in clinic is the typical, stoic...farmer that says "No, I've not had any help, I'm absolutely fine," he leaves the room and his wife goes "Yeah, he can't get to the toilet on his own" and this potentially picks up those problem patients because we do get those patients, not infrequently. And so this is a way of potentially flagging it up if it can do that.* [#0052]

Some participants expressed that the data could help avoid the common problem that reports by patients of function improvement are often not accurate as they are masked by pain:

> *I mean if you look at the Oxford Hip Score, 80% of the effects in the Oxford Hip Score are due to pain, so pain dominates in terms of what you see...That's why they tend to improve [be]cause you've reliably improved the pain...but this is the difference, this not reported function, this is real function and the two are quite different. "Can you go up and down the stairs?" is not the same as are you going up and down stairs? That's what's useful about this, isn't it?* [#0049]

However, a participant identified a dilemma if the home data contradict the patient's own account, potentially damaging the patient-clinician relationship:

> *You can't break that trust that you have to have, if someone says this is what I do, then I have to take that at face value, regardless of whether I believe it or not.* [#0046]

Data From the Home Provide a Stimulus for Discussion in a Consultation With a Patient

Although participants indicated that visualizations 4 to 8 were suitable only for use by clinicians, some visualizations were considered to be a good basis for discussion with patients:

> *The other thing that I think it would be really useful for is providing information to patients after the operation. So you could monitor a cohort of total hips, total knees, hip fracture, different patients and say "We can expect that your sleep will have returned to normal after eight weeks" or "You will be leaving*

XSL•FO
RenderX

> *the house more back the way you were at six weeks" and that would be really useful. So there are some things we get from patients and we tell them that we're told that on average you get back to bowls [note: the sport of lawn bowls] at six weeks, but actually having a bit of an evidence base to say "...most [patients] were sleeping through the night by six weeks." That would be a nice thing to be able to say to patients.* [#0052]

> *If they say, "I'm not sleeping well," and you still look and you say, well although you're not sleeping perfectly, you've definitely improved over the last six weeks—do you see what I mean? [Be]cause they don't always remember that.* [#0049]

Surgeons felt that it would be feasible for them to use a decision support system in consultations with patients. Given the inevitable complexity of data derived from people's daily lives, automated processing of data was preferred over a presentation of relatively raw data. Surgeons found the breadth of possible patient information fascinating. However, many said that the need for speed in their necessarily brief consultations left little room to conduct anything other than the "essentials":

> *It's got huge amount of potential, I just don't know what to do with all these lovely graphs and figures really.* [#0052]

> *I think it's fascinating to see and I think but the reality at the coalface is that in a clinical situation, you just need to do essentials as quickly as possible. I'm struggling to see how that could happen in the ordinary, everyday situation because in this early phase, patients' recovery trajectories will vary very much...during this early phase, depending on their co-morbidities and everything else, there's a very different speed of achieving certain milestones.* [#0051]

> *I do think that there is a time element there when you're using the data...if you had a summary page and that was compared to what a normal recovery would be, like a traffic light system. It's way too much information to process in a clinic.* [#0051]

> *Sadly I can't get past the fact for routine follow up of post-op [hip replacement] patients, we're already cutting back how many we see and what we do, because they all tend to do so well and so giving us more information is probably not helpful.* [#0050]

Use of Home Data for Research Purposes

A possibility was that the data could be developed into outcome measures for research purposes, with the overall aim to be able to consult such information when addressing individual patient cases:

> *Certainly from a research point of view if you're wanting to follow something up like a new hip prosthesis and you wanted to know whether this was making any difference in this early phase...this could be very useful in supporting that.* [#0051]

Using the data as an outcome measure for research was felt to have considerable potential:

> *I think the power of this is on a clinical basis, we could do more pilot stuff, you can correlate that with your interviews with the patients.* [#0046]

Meeting Surgeons' Requirements in the Development of Visualizations

Most participants (8/9, 89%) identified concerns regarding existing visualizations and proposed a way to address them. Challenges included the difficulty of representing large amounts of time-based data without losing detail, accessibility of visualizations to patients, the time required to interpret the visualization, and the provision of excessive detail. To address these issues, surgeons suggested that goal-focused visualizations that solve a small number of competency questions would be of value. For instance, charts showing "a simple form of data" could more easily support clinical evaluation of "one specific activity."

Discussion

Principal Findings

On the basis of an action research approach [41], this paper reports the findings of scenario-based focus groups and interviews. This study aimed to provide insights into the presentation of time-series data as a way of assessing recovery after surgery and to what extent the data supported clinical workflows. Participants generally noted that the data offered a more objective assessment of patient recovery than current methods used in their clinical practice.

Of the visualizations presented, a dashboard comprising specific notifications and alerts seemed to be the best fit for existing workflows. Automation of clinical decisions based on "moment-to-moment quantification of individual level data" [53] and rapidly condensing large amounts of data into meaningful information aligned with the 10-minute appointment time that NHS surgeons have with patients at follow-up.

The tabular and circular data visualizations spanning longer periods were considered useful by surgeons for identifying trends and changes ahead of the consultation. Furthermore, the granular detail of patients' movement trajectory immediately before and after surgery was considered useful within a consultation, in which the surgeon and patient could address expectations of outcomes after surgery and longer-term follow-up.

It was noted that such discussions would require assurance of sensitive and accurate interpretation of any data beforehand to avoid any negative impact of patient engagement with the data. Furthermore, it would be necessary to decide whether to measure the patient's progress in absolute terms with reference to a population mean or purely relative to their own initial baseline.

Surgeons are not accustomed to visualizing and conceptualizing time-series data from the home, and as with any new form of clinical data, undoubtedly training would be a prerequisite for the introduction of this type of data into clinical practice. The

XSL•FO
RenderX

participants in this study were interested in the complexity of the granular data and were aware that insights would be lost if they were summarized or averaged. However, they had not been provided with professional development training in interpretation of the data. It is reasonable to suppose that training and familiarity would unlock more of the value in the data and lessen some of the legitimate concerns about the data being confusing or time-consuming to use. The challenge of finding intuitive ways of presenting weeks of continuous data to clinicians for use in a 10-minute clinical appointment would be a good area for future research.

Comparison With Prior Work

The UK National Joint Registry recently introduced a patient decision support tool that aimed to enhance patients' understanding of their own risks and the potential benefits of having joint replacement surgery [54]. Innovative tools may empower patients to have informed conversations with their physicians about treatment options, and such tools can support evidence-based choices, moving closer toward personalized medicine. Our findings suggest that a clinical decision support system that tracks and interprets activity at home could supplement such information, further enhancing a patient's choices about treatment options and postrecovery options after surgery. Furthermore, by collecting data before and after surgery, there is the chance to compare outcomes after surgery with presurgical ability and help in communication about expectations before surgery and whether those expectations have been met.

Clinical decision support systems make use of appropriate data analytics and visualization methods to provide advice and guidance to aid health care providers' problem-solving and decision-making [55]. Potential benefits of designing clinical decision support systems include improving consistency in decision-making, increasing efficiency, and reducing task interruptions and the corresponding fatigue [56]. A recent randomized controlled trial that evaluated the use of a patient decision aid and preference report (ie, a summary of patient clinical and decisional data) by surgeons performing joint replacement [57] found that this supports shared decision-making between clinician and patient and that there was significant improvement in decision quality when such aids were used. The findings of our study are in accordance with the proposal of increased efficiency to some degree as surgeons thought that although some of the home data would be helpful, it would be unlikely that they would directly reference or share this information in consultation because of the limited time they had with patients.

The wide range of sensors available to patients as wearables or within smart home products can help patients track exercise, sleep, heart rate, and much more. Data collected by such sensors to improve health have been used to help with diagnosis and monitoring in the fields of chronic health conditions [58] and mental health [59]. This study illustrates the potential for home data of this kind to be used to support clinical follow-ups after hip and knee replacement surgery. Our study presents a novel exploration of movement data collected via a platform of sensors for use in orthopedics, aiming to yield new ways of advancing conventional follow-up assessments following total hip and knee replacement surgery.

Strengths

A key strength of this study was the use of an action research approach, which included an exploratory phase followed by discussion of a proposed model of data presentation using real patients' stories. The triangulation of patterns detected in the quantitative data with real patient participants' accounts from qualitative interviews contributed to a robust analysis of the data with a good degree of accuracy. A qualitative analysis of patients' experiences has been reported elsewhere [45]. Finally, this study uniquely explores surgeons' views of data visualization from novel sensing technology, which is not currently commercially available but could be put on the market in the near future if desired. Insights from this study can help inform research and design directions for products in this space.

Limitations

The sample comprised surgeons from 1 UK hospital and, as such, only reflects experience in 1 setting. Convenience sampling was used, which may limit the ability to generalize from this sample. However, in practice, participants' background, skills, and experience were heterogeneous, as were their age and sex. The experiences described are likely to be consonant with those in other contexts, and all UK surgeons follow national approaches and training. More importantly, there may be differences between the findings of our study and those that are relevant in other countries; although the surgical procedure is similar in different contexts, patients' expectations and the resources available to surgeons may vary. Furthermore, in the United Kingdom and internationally, professionals other than surgeons are involved in the provision of care to patients undergoing knee or hip replacement. For instance, specialist physiotherapists are involved in assessments before surgery and provide care afterward. We did not include their professional experience in this study, and this could be a topic of further research; however, in practice, most health professionals face similar challenges related to time pressures on consultations and the need to collect and convey clear and relevant information.

Everyday practice following the COVID-19 pandemic has required adjustment to deal with service backlogs, such as a move toward day case surgery as well as decreasing length of stay and adoption of remote assessment of postoperative recovery status. It is not yet clear to what extent what proportion of sites has moved to this model and what proportion of patients are affected by this change. There is also a move toward patient-initiated follow-up. However, this is at an early stage, and it remains to be seen what the benefits and shortcomings might be for patients and participants.

Conclusions and Recommendations

Overview

In line with an action research approach [41], we propose the following 4 guidelines for further design and development of home activity monitoring systems. Each guideline draws on the findings described previously and was codeveloped by the

XSL•FO
RenderX

interdisciplinary group of coauthors in light of the findings. As such, the guidelines consolidate the views of surgeons and the thematic areas developed in this study and provide recommendations for next steps, including how best to support surgeons—or other health care professionals—and how best to design and deliver a user-appropriate system. Each guideline reflects the content of more than one thematic area. Our aim was to build on surgeons' views to provide concrete recommendations to support future developments in the collection, visualization, and use of data on recovery or other health changes.

Guideline 1: Minimize the Risk of Misinterpreting Data

Surgeons demonstrated consensus on the importance of reducing the risk of misinterpretation of data and the associated variability of interpretations between surgeons. To minimize the risk of misinterpretation, clear summary statistics are recommended. Explainable design principles appropriate for each visualization or presentation of data should be applied to clarify the meaning and limitations of the data and the associated findings. It would be misguided to promise absolute objectivity as the activities of data acquisition, data analysis, and machine learning frequently result in the reproduction of bias present in source data or in the unconscious predispositions held by data analysts themselves [60].

Guideline 2: Express the Level of Confidence in the Data

Surgeons expressed a preference for simple and unambiguous metrics. However, electronics and sensor systems in the home inevitably experience many challenges to reliability, such as device failure or wireless network failures; therefore, data from such an uncontrolled environment must always be interpreted with caution, and a level of confidence would need to accompany any data analysis.

Guideline 3: Improve Familiarity With Time-Series Data

Exploratory methods of accessing big data are a poor fit with constraints on surgeons' time. Efficient, rapidly accessible representations of home data requiring minimal expert knowledge are recommended in the first instance. For example, data summarization can facilitate the interpretation of complex data, removing outliers and supporting existing clinical consultation activities. The 2019 Topol Review [61] indicates that training and digital literacy are key to making the most of digital health technologies, particularly artificial intelligence and machine learning. Identifying an understanding of confidence and probability is a necessary prerequisite for interpreting these data and is a required skill. We suggest that familiarity with time-series representations of data acquired through training may increasingly be an advantageous skill for clinical purposes.

Guideline 4: Consider the Impact of Patient Engagement

Data are of interest to surgeons as a resource that they can use to assist in their communications with patients. Future developments such as interfaces that support patients in examining their own data may offer a level of empowerment. Greater patient empowerment is positively associated with adherence to treatments and improved outcomes [62]. It also supports the UK NHS commitment to person-centered care, in which patients are encouraged to be actively involved in their own care [63]. Therefore, patient-centered design practices are a substantial component in the development of systems that use home data to support patient-clinician interactions. The time constraints experienced by surgeons limit their opportunities to have direct overview of time-series home data. A patient-centric approach could support patients in monitoring changes in their own condition, potentially facilitating conversations with clinicians. Finally, the schedule by which surgeons or other clinicians review data is not a close fit with the potential for "just-in-time" alerting systems, suggesting that some of the potential of home data may rely on structural innovation and integration with wider support teams.

Acknowledgments

This work was carried out under the Sensor Platform for Healthcare in a Residential Environment Interdisciplinary Research Collaboration, funded by the UK Engineering and Physical Sciences Research Council, grant EP/K031910/1.

Authors' Contributions

All authors contributed to the design of the study. MH, AM, MPN, HS, and ET developed the visualizations and provided explanatory material. SG, MH, and MPN conducted the participant interviews, and SG analyzed the interview data with support from AB, AJ, and RGH. SG, ET, and IC wrote the manuscript. All authors reviewed the final manuscript.

Conflicts of Interest

RGH is the co-chair of the UK Committee on Research Integrity. The authors have no further conflicts to declare.

Multimedia Appendix 1
Topic guides used during this study.
[PDF File (Adobe PDF File), 89 KB - periop_v6i1e36172_app1.pdf]

References

XSL•FO
RenderX

1. Northern IIOM. The National Joint Registry 16th Annual Report 2019 [Internet]. London, UK: National Joint Registry; 2019.
2. Maradit Kremers H, Larson DR, Crowson CS, Kremers WK, Washington RE, Steiner CA, et al. Prevalence of total hip and knee replacement in the United States. J Bone Joint Surg Am 2015 Sep 02;97(17):1386-1397 [FREE Full text] [doi: 10.2106/JBJS.N.01141] [Medline: 26333733]
3. Cross M, Smith E, Hoy D, Nolte S, Ackerman I, Fransen M, et al. The global burden of hip and knee osteoarthritis: estimates from the global burden of disease 2010 study. Ann Rheum Dis 2014 Jul;73(7):1323-1330. [doi: 10.1136/annrheumdis-2013-204763] [Medline: 24553908]
4. Joint replacement (primary): hip, knee and shoulder. National Institute for Health and Care Research. 2020 Jun 4. URL: https://www.nice.org.uk/guidance/ng157 [accessed 2021-11-10]
5. Patient reported outcome measures (PROMs). NHS England. URL: https://www.england.nhs.uk/statistics/statistical-work-areas/proms/ [accessed 2017-02-26]
6. Ware J, Kosinski M, Keller SD. A 12-item short-form health survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996 Mar;34(3):220-233. [doi: 10.1097/00005650-199603000-00003] [Medline: 8628042]
7. Janssen MF, Pickard AS, Golicki D, Gudex C, Niewada M, Scalone L, et al. Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. Qual Life Res 2013 Sep;22(7):1717-1727 [FREE Full text] [doi: 10.1007/s11136-012-0322-4] [Medline: 23184421]
8. Guyatt GH. A taxonomy of health status instruments. J Rheumatol 1995 Jun;22(6):1188-1190. [Medline: 7674253]
9. Dawson J, Fitzpatrick R, Carr A, Murray D. Questionnaire on the perceptions of patients about total hip replacement. J Bone Joint Surg Br 1996 Mar;78(2):185-190. [Medline: 8666621]
10. Dawson J, Fitzpatrick R, Murray D, Carr A. Questionnaire on the perceptions of patients about total knee replacement. J Bone Joint Surg Br 1998 Jan;80(1):63-69. [doi: 10.1302/0301-620x.80b1.7859] [Medline: 9460955]
11. Wilson I, Bohm E, Lübbeke A, Lyman S, Overgaard S, Rolfson O, et al. Orthopaedic registries with patient-reported outcome measures. EFORT Open Rev 2019 Jun;4(6):357-367 [FREE Full text] [doi: 10.1302/2058-5241.4.180080] [Medline: 31210973]
12. Field J, Holmes MM, Newell D. PROMs data: can it be used to make decisions for individual patients? A narrative review. Patient Relat Outcome Meas 2019 Jul;Volume 10:233-241. [doi: 10.2147/prom.s156291]
13. Zini ML, Banfi G. A narrative literature review of bias in collecting patient reported outcomes measures (PROMs). Int J Environ Res Public Health 2021 Nov 26;18(23):12445 [FREE Full text] [doi: 10.3390/ijerph182312445] [Medline: 34886170]
14. Ostendorf M, van Stel HF, Buskens E, Schrijvers AJ, Marting LN, Verbout AJ, et al. Patient-reported outcome in total hip replacement. A comparison of five instruments of health status. J Bone Joint Surg Br 2004 Aug;86(6):801-808. [doi: 10.1302/0301-620x.86b6.14950] [Medline: 15330018]
15. Roos EM, Lohmander LS. The knee injury and osteoarthritis outcome score (KOOS): from joint injury to osteoarthritis. Health Qual Life Outcome 2003 Nov 03;1:64 [FREE Full text] [doi: 10.1186/1477-7525-1-64] [Medline: 14613558]
16. Kuhn M, Harris-Hayes M, Steger-May K, Pashos G, Clohisy JC. Total hip arthroplasty in patients 50 years or less: do we improve activity profiles? J Arthroplasty 2013 May;28(5):872-876. [doi: 10.1016/j.arth.2012.10.009] [Medline: 23499406]
17. Bolink SA, Lenguerrand E, Brunton LR, Wylde V, Gooberman-Hill R, Heyligers IC, et al. Assessment of physical function following total hip arthroplasty: inertial sensor based gait analysis is supplementary to patient-reported outcome measures. Clin Biomech (Bristol, Avon) 2016 Feb;32:171-179. [doi: 10.1016/j.clinbiomech.2015.11.014] [Medline: 26706048]
18. Diethe T, Holmes M, Kull M, Perello NM, Sokol K, Song H, et al. Releasing ehealth analytics into the wild: lessons learnt from the SPHERE project. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2018 Presented at: KDD '18: The 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; Aug 19 - 23, 2018; London, United Kingdom. [doi: 10.1145/3219819.3219883]
19. Wallis JA, Webster KE, Levinger P, Taylor NF. What proportion of people with hip and knee osteoarthritis meet physical activity guidelines? A systematic review and meta-analysis. Osteoarthritis Cartilage 2013 Nov;21(11):1648-1659 [FREE Full text] [doi: 10.1016/j.joca.2013.08.003] [Medline: 23948979]
20. Grimm B, Bolink S. Evaluating physical function and activity in the elderly patient using wearable motion sensors. EFORT Open Rev 2016 May;1(5):112-120 [FREE Full text] [doi: 10.1302/2058-5241.1.160022] [Medline: 28461937]
21. Kangas M, Korpelainen R, Vikman I, Nyberg L, Jämsä T. Sensitivity and false alarm rate of a fall sensor in long-term fall detection in the elderly. Gerontology 2015 Aug 13;61(1):61-68. [doi: 10.1159/000362720] [Medline: 25138139]
22. Woznowski P, Burrows A, Diethe T, Fafoutis X, Hall J, Hannuna S, et al. SPHERE: a sensor platform for healthcare in a residential environment. In: Designing, Developing, and Facilitating Smart Cities. Cham: Springer; 2017.
23. Elsts A, Fafoutis X, Woznowski P, Tonkin E, Oikonomou G, Piechocki R, et al. Enabling healthcare in smart homes: the SPHERE IoT network infrastructure. IEEE Commun Mag 2018 Dec;56(12):164-170. [doi: 10.1109/mcom.2017.1700791]
24. Holmes M, Nieto MP, Song H, Tonkin E, Grant S, Flach P. Modelling patient behaviour using IoT sensor data: a case study to evaluate techniques for modelling domestic behaviour in recovery from total hip replacement surgery. J Healthc Inform Res 2020 Sep 03;4(3):238-260 [FREE Full text] [doi: 10.1007/s41666-020-00072-6] [Medline: 35415449]
25. Masullo A, Burghardt T, Perrett T, Damen D, Mirmehdi M. Sit-to-stand analysis in the wild using silhouettes for longitudinal health monitoring. In: Image Analysis and Recognition. Cham: Springer; 2019.

XSL•FO
RenderX

26. Anders S, Albert R, Miller A, Weinger MB, Doig AK, Behrens M, et al. Evaluation of an integrated graphical display to promote acute change detection in ICU patients. Int J Med Inform 2012 Dec;81(12):842-851 [FREE Full text] [doi: 10.1016/j.ijmedinf.2012.04.004] [Medline: 22534099]
27. Stubbs B, Kale DC, Das A. Sim•TwentyFive: an interactive visualization system for data-driven decision support. AMIA Annu Symp Proc 2012;2012:891-900 [FREE Full text] [Medline: 23304364]
28. Martinez-Garcia A, Naranjo-Saucedo AB, Rivas JA, Romero Tabares A, Marín Cassinello A, Andrés-Martín A, et al. A clinical decision support system (KNOWBED) to integrate scientific knowledge at the bedside: development and evaluation study. JMIR Med Inform 2021 Mar 10;9(3):e13182 [FREE Full text] [doi: 10.2196/13182] [Medline: 33709932]
29. Silveira DV, Marcolino MS, Machado EL, Ferreira CG, Alkmim MB, Resende ES, et al. Development and evaluation of a mobile decision support system for hypertension management in the primary care setting in Brazil: mixed-methods field study on usability, feasibility, and utility. JMIR Mhealth Uhealth 2019 Mar 25;7(3):e9869 [FREE Full text] [doi: 10.2196/mhealth.9869] [Medline: 30907740]
30. Cao J, Truong AL, Banu S, Shah AA, Sabharwal A, Moukaddam N. Tracking and predicting depressive symptoms of adolescents using smartphone-based self-reports, parental evaluations, and passive phone sensor data: development and usability study. JMIR Ment Health 2020 Jan 24;7(1):e14045 [FREE Full text] [doi: 10.2196/14045] [Medline: 32012072]
31. Gandrup J, Yazdany J. Using health information technology to support use of patient-reported outcomes in rheumatology. Rheum Dis Clin North Am 2019 May;45(2):257-273 [FREE Full text] [doi: 10.1016/j.rdc.2019.01.007] [Medline: 30952397]
32. Qiu S, Liu L, Wang Z, Li S, Zhao H, Wang J, et al. Body sensor network-based gait quality assessment for clinical decision-support via multi-sensor fusion. IEEE Access 2019;7:59884-59894. [doi: 10.1109/access.2019.2913897]
33. Billis A, Katzouris N, Artikis A, Bamidis P. Clinical decision support for active and healthy ageing: an intelligent monitoring approach of daily living activities. In: Progress in Artificial Intelligence. Cham: Springer; 2015:2015.
34. De Rosis S, Pennucci F, Lungu DA, Manca M, Nuti S. A continuous PREMs and PROMs observatory for elective hip and knee arthroplasty: study protocol. BMJ Open 2021 Sep 21;11(9):e049826 [FREE Full text] [doi: 10.1136/bmjopen-2021-049826] [Medline: 34548358]
35. Lamberigts M, Van Hoof L, Proesmans T, Vandervoort P, Grieten L, Haemers P, et al. Remote heart rhythm monitoring by photoplethysmography-based smartphone technology after cardiac surgery: prospective observational study. JMIR Mhealth Uhealth 2021 Apr 15;9(4):e26519 [FREE Full text] [doi: 10.2196/26519] [Medline: 33856357]
36. Backonja U, Haynes SC, Kim KK. Data visualizations to support health practitioners' provision of personalized care for patients with cancer and multiple chronic conditions: user-centered design study. JMIR Hum Factors 2018 Oct 16;5(4):e11826 [FREE Full text] [doi: 10.2196/11826] [Medline: 30327290]
37. Barbour R, Kitzinger J, editors. Developing Focus Group Research Politics, Theory and Practice. Thousand Oaks, California, United States: SAGE Publications; 1998.
38. Malterud K, Siersma VD, Guassora AD. Sample size in qualitative interview studies: guided by information power. Qual Health Res 2016 Nov;26(13):1753-1760. [doi: 10.1177/1049732315617444] [Medline: 26613970]
39. Grant S, Blom AW, Whitehouse MR, Craddock I, Judge A, Tonkin EL, et al. Using home sensing technology to assess outcome and recovery after hip and knee replacement in the UK: the HEmiSPHERE study protocol. BMJ Open 2018 Jul 28;8(7):e021862 [FREE Full text] [doi: 10.1136/bmjopen-2018-021862] [Medline: 30056388]
40. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh sleep quality index: a new instrument for psychiatric practice and research. Psychiatry Res 1989 May;28(2):193-213. [doi: 10.1016/0165-1781(89)90047-4] [Medline: 2748771]
41. Elg M, Engström J, Witell L, Poksinska B. Co-creation and learning in health-care service development. J Serv Manage 2012;23(3):328-343. [doi: 10.1108/09564231211248435]
42. Carroll JM. Making Use Scenario-Based Design of Human-Computer Interactions. Cambridge, Massachusetts, United States: MIT Press; 2000.
43. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
44. Ritchie J, Lewis J, Nicholls CM, Ormston R, editors. Qualitative Research Practice A Guide for Social Science Students and Researchers. Thousand Oaks, California, United States: SAGE Publications; 2013.
45. Grant S, Blom AW, Craddock I, Whitehouse M, Gooberman-Hill R. Home health monitoring around the time of surgery: qualitative study of patients' experiences before and after joint replacement. BMJ Open 2019 Dec 15;9(12):e032205 [FREE Full text] [doi: 10.1136/bmjopen-2019-032205] [Medline: 31843836]
46. Carter N, Bryant-Lukosius D, DiCenso A, Blythe J, Neville AJ. The use of triangulation in qualitative research. Oncol Nurs Forum 2014 Sep;41(5):545-547. [doi: 10.1188/14.ONF.545-547] [Medline: 25158659]
47. Bakrania K, Yates T, Rowlands AV, Esliger DW, Bunnewell S, Sanders J, et al. Intensity thresholds on raw acceleration data: euclidean norm minus one (ENMO) and mean amplitude deviation (MAD) approaches. PLoS One 2016 Oct 5;11(10):e0164045 [FREE Full text] [doi: 10.1371/journal.pone.0164045] [Medline: 27706241]
48. Heer J, Kong N, Agrawala M. Sizing the horizon: the effects of chart size and layering on the graphical perception of time series visualizations. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2009 Presented

XSL•FO
RenderX

at: CHI '09: CHI Conference on Human Factors in Computing Systems; Apr 4 - 9, 2009; Boston, MA, USA. [doi: 10.1145/1518701.1518897]
49. Weber M, Alexa M, Müller W. Visualizing time-series on spirals. In: Proceedings of the IEEE Symposium on Information Visualization, 2001. INFOVIS 2001. 2001 Presented at: IEEE Symposium on Information Visualization, 2001. INFOVIS 2001; Oct 22-23 , 2001; San Diego, CA, USA. [doi: 10.1109/infvis.2001.963273]
50. Holmes M, Nieto MP, Song H, Tonkin E, Grant S, Flach P. Analysis of patient domestic activity in recovery from Hip or Knee replacement surgery: modelling wrist-worn wearable RSSI and Accelerometer data in the wild. In: Proceedings of the 3rd International Workshop on Knowledge Discovery in Healthcare Data @IJCAI. 2018 Presented at: The 3rd International Workshop on Knowledge Discovery in Healthcare Data @IJCAI; Jul, 2018; Stockholm, Sweden.
51. Janssen WG, Bussmann HB, Stam HJ. Determinants of the sit-to-stand movement: a review. Phys Ther 2002 Sep;82(9):866-879. [Medline: 12201801]
52. Miller K, Capan M, Weldon D, Noaiseh Y, Kowalski R, Kraft R, et al. The design of decisions: matching clinical decision support recommendations to Nielsen's design heuristics. Int J Med Inform 2018 Sep;117:19-25 [FREE Full text] [doi: 10.1016/j.ijmedinf.2018.05.008] [Medline: 30032961]
53. Vaidyam A, Halamka J, Torous J. Actionable digital phenotyping: a framework for the delivery of just-in-time and longitudinal interventions in clinical healthcare. Mhealth 2019;5:25 [FREE Full text] [doi: 10.21037/mhealth.2019.07.04] [Medline: 31559270]
54. The National Joint Registry homepage. The National Joint Registry. URL: https://www.njrcentre.org.uk/ [accessed 2021-11-10]
55. Greenes R, editor. Clinical Decision Support The Road Ahead. London: Elsevier Science; 2011.
56. Sittig DF, Wright A, Osheroff JA, Middleton B, Teich JM, Ash JS, et al. Grand challenges in clinical decision support. J Biomed Inform 2008 Apr;41(2):387-392 [FREE Full text] [doi: 10.1016/j.jbi.2007.09.003] [Medline: 18029232]
57. Stacey D, Hawker G, Dervin G, Tugwell P, Boland L, Pomey M, et al. Decision aid for patients considering total knee arthroplasty with preference report for surgeons: a pilot randomized controlled trial. BMC Musculoskelet Disord 2014 Feb 24;15(1):54 [FREE Full text] [doi: 10.1186/1471-2474-15-54] [Medline: 24564877]
58. Vegesna A, Tran M, Angelaccio M, Arcona S. Remote patient monitoring via non-invasive digital technologies: a systematic review. Telemed J E Health 2017 Jan;23(1):3-17 [FREE Full text] [doi: 10.1089/tmj.2016.0051] [Medline: 27116181]
59. Cornet VP, Holden RJ. Systematic review of smartphone-based passive sensing for health and wellbeing. J Biomed Inform 2018 Jan;77:120-132 [FREE Full text] [doi: 10.1016/j.jbi.2017.12.008] [Medline: 29248628]
60. Brodie MA, Pliner EM, Ho A, Li K, Chen Z, Gandevia SC, et al. Big data vs accurate data in health research: large-scale physical activity monitoring, smartphones, wearable devices and risk of unconscious bias. Med Hypotheses 2018 Oct;119:32-36. [doi: 10.1016/j.mehy.2018.07.015] [Medline: 30122488]
61. Preparing the healthcare workforce to deliver the digital future. Health Education England. 2019. URL: https://topol.hee.nhs.uk/wp-content/uploads/HEE-Topol-Review-2019.pdf [accessed 2021-11-10]
62. Chan DK, Lonsdale C, Ho PY, Yung PS, Chan KM. Patient motivation and adherence to postsurgery rehabilitation exercise recommendations: the influence of physiotherapists' autonomy-supportive behaviors. Arch Phys Med Rehabil 2009 Dec;90(12):1977-1982. [doi: 10.1016/j.apmr.2009.05.024] [Medline: 19969157]
63. Next steps on the NHS Five Year Forward View. NHS. 2019. URL: https://www.england.nhs.uk/wp-content/uploads/2017/03/NEXT-STEPS-ON-THE-NHS-FIVE-YEAR-FORWARD-VIEW.pdf [accessed 2021-11-10]

Abbreviations

NHS: National Health Service
OHS: Oxford Hip Score
OKS: Oxford Knee Score
PROM: patient-reported outcome measure
SPHERE: Sensor Platform for Healthcare in a Residential Environment

XSL•FO
RenderX

Edited by T Leung; submitted 07.01.22; peer-reviewed by AM Vranceanu, J Egger; comments to author 03.05.22; revised version received 14.11.22; accepted 16.02.23; published 24.04.23.

Please cite as:
Grant S, Tonkin E, Craddock I, Blom A, Holmes M, Judge A, Masullo A, Perello Nieto M, Song H, Whitehouse M, Flach P, Gooberman-Hill R
Toward Enhanced Clinical Decision Support for Patients Undergoing a Hip or Knee Replacement: Focus Group and Interview Study With Surgeons
JMIR Perioper Med 2023;6:e36172
URL: https://periop.jmir.org/2023/1/e36172
doi:10.2196/36172
PMID:37093626

©Sabrina Grant, Emma Tonkin, Ian Craddock, Ashley Blom, Michael Holmes, Andrew Judge, Alessandro Masullo, Miquel Perello Nieto, Hao Song, Michael Whitehouse, Peter Flach, Rachael Gooberman-Hill. Originally published in JMIR Perioperative Medicine (http://periop.jmir.org), 24.04.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Perioperative Medicine, is properly cited. The complete bibliographic information, a link to the original publication on http://periop.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

Patient Safety of Perioperative Medication Through the Lens of Digital Health and Artificial Intelligence

Jiancheng Ye[1], PhD

Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

Corresponding Author:
Jiancheng Ye, PhD
Feinberg School of Medicine
Northwestern University
633 N Saint Clair St
Chicago, IL, 60611
United States
Phone: 1 312 503 3690
Email: jiancheng.ye@u.northwestern.edu

Abstract

Perioperative medication has made significant contributions to enhancing patient safety. Nevertheless, administering medication during this period still poses considerable safety concerns, with many errors being detected only after causing significant physiological disturbances. The intricacy of medication administration in the perioperative setting poses specific challenges to patient safety. To address these challenges, implementing potential strategies and interventions is critical. One such strategy is raising awareness and revising educational curricula regarding drug safety in the operating room. Another crucial strategy is recognizing the importance of redundancy and multiple checks in the operating room as a hallmark of medication safety, which is not a common practice. Digital health technologies and artificial intelligence (AI) also offer the potential to improve perioperative medication safety. Computerized physician order entry systems, electronic medication administration records, and barcode medication administration systems have been proven to reduce medication errors and improve patient safety. By implementing these strategies and interventions, health care professionals can enhance the safety of perioperative medication administration and improve patient outcomes.

(JMIR Perioper Med 2023;6:e34453) doi:10.2196/34453

KEYWORDS

perioperative medicine; patient safety; anesthesiology; human factors; medication errors; digital health; health information technology

Introduction

Medication errors represent a critical patient safety problem, arising from failures in completing required actions or using the wrong plan or action to achieve patient care aims [1]. These errors can be classified by type, including incorrect dose, substitution, omission, repetition, insertion, and unattended drug use [2]. In the perioperative setting, the administration of intravenous (IV) medications, or the “medication use process” [3], presents unique patient safety obstacles. The medication use process consists of several steps, including requesting, dispensing, preparing, administering, documenting, and monitoring patients for the effects of medication [4]. However, compared to almost any other hospital setting, medication administration in the operating room lacks most standard and accepted safety checks [1]. Unlike nurses, who require a physician to place an order for a medication, the pharmacy to prepare the medication, and a second nurse to verify the medication prior to administration, the anesthesiologist working alone in the operating room can determine the need for a medication (ie, diagnose and prescribe), draw up the medication (ie, prepare), administer and monitor the effects of the medication, and record events without any verification check for safety and accuracy [5]. Furthermore, the often fast-paced and high-stress environment of the operating room can further increase the likelihood of medication errors. Thus, it is not surprising that medication errors are common in this setting [6]. To address this issue, it is essential to implement strategies and interventions that improve the safety and accuracy of the process of medication use in the operating room.

XSL•FO
RenderX

Patient Safety Problems Associated With Perioperative Medication Errors

Medication errors can be classified based on their potential for patient harm and whether they result in an adverse drug event or not. These classifications include errors with no potential for harm (near miss), those with little potential for patient harm, those with potential for adverse drug events, and those resulting in adverse drug events [1]. An adverse drug event is defined as any patient injury resulting from medication [7]. However, it is essential to note that adverse drug events can occur even without medication errors; for example, in the case of an allergic reaction. Nanji et al [1] further classified medication errors and adverse drug events by their severity (significant, serious, life-threatening, and fatal) and preventability (definitely preventable, probably preventable, probably not preventable, and definitely not preventable), and they found that out of the 193 (of a total of 3671, 5.3%) identified medication errors and adverse drug events, 153 (79.3%) were preventable. Additionally, 32 (20.9%) of these medication errors had little potential for harm, 70 (45.8%) had the potential for patient harm, and 51 (33.3%) resulted in an adverse drug event. The errors were further classified as serious (n=99, 64.7%), significant (n=51, 33.3%), and life-threatening (n=3, 2%), with no fatalities attributable to medication errors. In a separate study by Cooper et al [2], the authors identified 52 medication errors, resulting in no harm in 24 patients, minor harm in 15 patients, and harm in 13 patients.

Risk Factors

Medication errors have been a concern since the 1970s, but the exact prevalence of these errors is still unknown due to underreporting [3]. According to Cooper et al [8], human factors are the primary cause of medication errors. These factors include failure to check, poor labeling, syringe swaps, decreased vigilance, fatigue, and production pressure [9]. Distraction, pressure to proceed, and misread labels were found to be the top 3 factors contributing to medication errors. Nanji et al [1] identified the following as the 3 most common medication errors: wrong dose, improper labeling, and failure to deliver the appropriate medication. High-acuity medications such as propofol, phenylephrine, and fentanyl were the most common perioperative medications involved in these errors [10].

Despite increased awareness and emphasis on perioperative medication safety by the Anesthesia Patient Safety Foundation, little progress has been made in addressing the human factor constraints that lead to medication errors. These constraints include lack of standardized labels; varied drug vial sizes, shapes, and colors; poorly designed medication carts and drug dispensing machines; and look-alike and sound-alike medications. The labeling and packaging of medications were found to be contributors to almost one-third of voluntarily reported medication errors leading to fatalities in the 1990s, according to the Institute of Safe Medication Practices' Medication Errors Reporting Program [11]. The Food and Drug Administration estimates that suboptimal labeling and packaging contribute to approximately 20% of medication errors. Moreover, drug shortages may force pharmacies to source unfamiliar substitute medications that can lead to errors, as reported by 21% of hospital pharmacists in a 2017 survey [12].

The high rate of medication errors in perioperative medication administration can be attributed to several factors, but perhaps the most significant is the anesthesia culture itself. The norm of an anesthesiologist performing the medication administration process with no oversight is deeply ingrained in professional and organizational culture. Prielipp et al [13] discussed the concept of "normalized deviance" in anesthesia practice, where departure from correct behavior becomes so ingrained in work culture that it is no longer considered deviant. Despite evidence that independent double checks can detect up to 95% of potential medication errors and eliminate 58% of those identified [14], the culture of autonomy, rejection of "cookbook" medicine, and resistance to standardization hinder the reporting of minor or near-miss events and impede efforts to improve medication safety in anesthesia. Grigg and Roesler [3] summarized this culture as allowing providers to "hand scrawl on poorly labeled syringes drawn up from nonstandard, look-alike vials in a distracting environment and organize them in an arbitrary, personalized arrangement." Organizational barriers also exist, such as fear of reprisal for errors or a blame culture, which can prevent medication errors from being reported [15].

Interventions and Recommendations

Inadequate labeling and packaging of medications is one of the primary human factors associated with a significant number of medication errors. Poor labeling can result in the administration of the wrong medication, incorrect dosage, and an incorrect route of administration. Common labeling errors include syringes that are unlabeled or inaccurately labeled and illegible, handwritten labels [16]. To address the labeling problem, various solutions have been proposed. These include using standardized color-coded labels for similar drug classes, implementing barcode-assisted labeling systems to generate medication labels, and using commercially prepared prefilled medication syringes. In a systematic review by Maximous et al [17], improved labeling led to a reduction of 37% in medication errors. Recent data have questioned the effectiveness of color-coding in preventing medication errors [18], but research on human factors engineering (HFE) highlights the importance of pattern recognition when performing pressured, high-stress tasks, such as administering high-risk medication in the operating room environment [8]. Although support for the color-coded labeling system has somewhat decreased, HFE acknowledges that identifying objects in high-stress situations relies on multiple cues [3], and the historical use of color-coding remains a crucial cue for anesthesiologists.

Digital health technologies such as computerized physician order entry (CPOE) systems and electronic medication administration records (eMAR) have been shown to reduce medication errors and improve patient safety. CPOE systems allow health care professionals to electronically enter medication orders, reducing the likelihood of errors due to illegible handwriting or transcription errors. The use of CPOE systems can ensure that medication orders are accurately and efficiently transmitted to the pharmacy and the surgical team. The eMAR systems can provide real-time information about medication

XSL•FO
RenderX

administration during surgery, allowing the surgical team to make informed decisions about the patient's care [19]. eMAR reduces the risk of medication errors due to incorrect dosages or administration times. In the perioperative setting, these digital health technologies can be particularly beneficial.

Barcode medication administration (BCMA) technology has been successful on nursing floors, but is not yet widely implemented in anesthesia due to the cost of the systems and a lack of a universal electronic health record (EHR) system capable of scanning barcodes and incorporating the information into the operative record [20]. BCMA systems use barcodes to verify patient and medication information, thus reducing the risk of errors due to incorrect medication selection or administration. BCMA systems can help ensure that the correct medications are administered to the correct patient at the correct time, reducing the risk of medication errors and adverse events [21]. Point-of-care barcode scanning has the potential to eliminate 17% of medication errors and 25.5% of potential adverse drug events [1]. In a study by Merry et al [22], a 21% reduction in perioperative medication errors was demonstrated when syringe labels were scanned immediately before administration. When automated drug-specific decision support and alerts were added, an additional 29% and 59% of medication errors could be eliminated, respectively. Despite the availability of barcode-assisted labeling systems in their study environment, the authors found that up to 24% of medication errors still involved a labeling error. These errors occurred when health care professionals bypassed the system or found a workaround. To minimize these risks, prefilled medication syringes prepared at standard concentrations and provided by the pharmacy may be the best risk reduction strategy. A failure modes and effects analysis of the use of prefilled syringes has the potential to eliminate 16 medication preparation steps and 19 potential failure modes [23]. For organizations with barcode scanning abilities, these prefilled syringes could also use barcode technology, which would eliminate compounding of medications by the anesthesia clinicians and the associated risks of this practice. Other process-based interventions, such as facilitating timing of documentation, reducing workarounds, and standardizing connections of IV drug infusions to the most proximal port, could further reduce medication errors. These process interventions have the potential to eliminate 35%, 24%, and 1.3% of medication errors, respectively [1]. Ultimately, multimodal strategies are needed, which include all potential human factor system changes and process interventions discussed above [24]. Multimodal interventions, including barcode readers with automatic auditory and visual verification of the drug, prefilled color-coded syringes, and workspace improvements including standardized stocking of anesthesia carts, have the greatest potential to reduce errors [25]. In combination, these interventions could reduce error rates by 21%-35% per administration and 37%-41% per anesthetic [17].

Artificial intelligence (AI) has the potential to revolutionize medication safety by providing real-time decision support, reducing medication errors, and improving communication among health care professionals [26]. AI can analyze large data sets to identify patterns and predict adverse events, allowing health care professionals to intervene before harm occurs. AI can also provide decision support, suggesting the most appropriate medication and dosage for a particular patient based on their medical history and other factors. Finally, AI can facilitate communication among health care professionals, ensuring that critical information is shared in a timely and efficient manner. In the perioperative setting, AI can be particularly valuable in predicting and preventing adverse events. For example, AI algorithms can analyze vital signs and other patient data to identify patients at risk for postoperative complications such as sepsis or acute kidney injury [27]. AI can also provide decision support to help health care professionals select the most appropriate medication and dosage for a particular patient, taking into account their medical history, allergies, and other factors. Finally, AI can facilitate communication among health care professionals, ensuring that critical information is shared in a timely and efficient manner.

AI algorithms can analyze large amounts of data to identify patterns and predict medication errors. For example, AI can analyze medication orders and patient data to identify patients at high risk for medication errors. This can help clinicians to proactively intervene to prevent errors before they occur. Natural language processing algorithms can analyze free-text notes in the EHR to identify potential medication errors. For example, natural language processing can identify notes that mention medication errors or adverse drug events. This can help clinicians to identify and address medication errors that may have been missed through other means. AI can be used to provide real-time decision support to clinicians. For example, AI algorithms can analyze medication orders and provide alerts to clinicians about potential drug interactions, dosing errors, or other safety concerns. This can help clinicians to make informed decisions about medication orders and reduce the risk of errors. Table 1 outlines specific ways that AI can be used to improve perioperative medication errors.

XSL•FO
RenderX

Table 1. Application of artificial intelligence (AI) in the patient safety of perioperative medication.

AI technology	Applications
Predictive analytics	AI algorithms have the potential to analyze vast amounts of data, identify patterns, and predict medication errors [28]. For example, clinicians can proactively intervene to prevent errors before they occur by analyzing medication orders and patient data to identify patients at high risk for medication errors, using AI.
Natural language processing	Natural language processing algorithms can analyze free-text notes in the electronic health record to identify potential medication errors that may have been missed through other means, such as notes that mention medication errors or adverse drug events [29]. This approach can assist clinicians in identifying and addressing potential errors before they cause harm.
Clinical decision support	AI can provide real-time decision support to clinicians by analyzing medication orders and providing alerts for potential drug interactions, dosing errors, or other safety concerns [30]. This feature can assist clinicians in making informed decisions about medication orders and reducing the risk of errors.
Machine learning	Machine learning algorithms can be used to identify patterns and predict medication errors by analyzing medication orders and patient data. These algorithms can also be used to develop personalized medication regimens for individual patients based on their unique characteristics, which can improve medication safety and reduce the risk of adverse drug events [31].
Computer vision	Computer vision algorithms can be used with barcoding systems to verify medication administration [22]. For instance, computer vision can analyze barcode scans to verify that the medication matches the order and the patient's information in the electronic health record. This feature can help reduce the risk of errors due to incorrect medication administration.

Discussion

Ensuring patient safety is a paramount concern, especially when it comes to administering medications in the perioperative setting. Medication reconciliation is a crucial process that involves comparing a patient's medication orders with their current medication regimen to identify any discrepancies and prevent medication errors caused by incomplete or inaccurate medication histories. To ensure medication safety, it is important to perform preoperative medication reconciliation and document it accurately [32]. Standardized protocols for medication administration, such as those recommended by professional societies or institutions, can help reduce the risk of errors during drug preparation, dosage calculation, and administration [33]. Education and training are essential for improving the safety of perioperative medication. Clinicians should be trained on medication safety best practices, including the use of decision support tools, the importance of medication reconciliation, and the use of standardized protocols [34]. Effective communication among health care professionals, especially during handovers, is critical to reducing medication errors. The use of standardized communication tools and training can help improve communication among health care professionals [35]. Continuous electronic monitoring of vital signs, particularly during surgery, can help identify and manage medication-related adverse events promptly. Emerging evidence suggests that incorporating HFE principles into practice may improve patient safety by reducing cognitive workload and simplifying medication administration processes [36]. By optimizing the design of medication administration processes, we can reduce the risk of errors and improve patient safety [37].

Digital health technologies and AI can be used to enhance perioperative medication safety by detecting potential medication errors and providing decision support tools to clinicians in real time. Clinical decision support tools, such as alerts for potential drug interactions or incorrect doses, can help to prevent medication errors [38]. These tools can be integrated into EHR systems, providing clinicians with accurate and up-to-date information about medications and their effects. CPOE systems allow clinicians to enter medication orders directly into an EHR system, reducing errors caused by illegible handwriting or transcription errors [39]. These systems can also provide decision support tools, such as alerts for potential drug interactions or incorrect doses. Additionally, CPOE systems can help to standardize medication orders, reducing the risk of errors due to miscommunication or confusion. The eMAR systems allow health care professionals to electronically record medication administration, reducing errors due to incorrect dosages or administration times [40]. These systems can also be integrated with barcode scanning technology to ensure that the correct medication is administered to the correct patient at the correct time. Telepharmacy services can be used to provide medication-related support to health care professionals in remote or underserved areas, ensuring that medication orders are accurate and complete and that medications are administered safely [41]. Patient portals can be used to provide patients with information about their medications, including dosages, side effects, and potential interactions [42]. These portals can also be used to remind patients to take their medications and to provide instructions on how to properly administer medications. By incorporating these digital health technologies and AI, health care professionals can reduce the risk of medication errors and improve patient outcomes.

Improving patient safety during perioperative medication administration requires a multifaceted approach, incorporating strategies that address medication reconciliation, standardized protocols, effective communication, continuous monitoring, and HFE principles.

Conclusions

Perioperative medication safety has been largely overlooked in terms of rigorous assessment of medication events and the implementation of safety measures. Unlike other high-risk

XSL•FO
RenderX

industries, there are few safety protocols to prevent simple but dangerous medication errors, such as those related to labeling. Addressing systemic weaknesses that contribute to medication errors requires HFE and cultural reforms, and a shift away from focusing on individual blame and failure of truth-telling and transparency to enable real reform. Simply improving vigilance is insufficient since it does not address human factors and systemic issues that contribute to errors. Digital health technologies and AI offer significant promise in enhancing perioperative medication safety. Systems such as CPOE, eMAR, and BCMA can reduce medication errors and improve communication among health care professionals. AI can provide real-time decision support, predict adverse events, and facilitate communication. However, it is necessary to develop effective ways to measure medication errors and capture data to identify the true scope of the problem and develop solutions for mitigation. Standardization, medication reconciliation, education and training, clinical decision support, barcoding and electronic medication administration, and effective team communication are all crucial to improving perioperative medication safety. By implementing these strategies, health care professionals can reduce the risk of medication errors and improve patient safety.

Authors' Contributions

JY was responsible for the study design, data analysis, interpretation of the results, and writing of the manuscript.

Conflicts of Interest

None declared.

References

1. Nanji K, Patel A, Shaikh S, Seger DL, Bates DW. Evaluation of perioperative medication errors and adverse drug events. Anesthesiology 2016 Jan;124(1):25-34 [FREE Full text] [doi: 10.1097/ALN.0000000000000904] [Medline: 26501385]
2. Cooper L, DiGiovanni N, Schultz L, Taylor AM, Nossaman B. Influences observed on incidence and reporting of medication errors in anesthesia. Can J Anaesth 2012 Jun 31;59(6):562-570. [doi: 10.1007/s12630-012-9696-6] [Medline: 22467066]
3. Grigg EB, Roesler A. Anesthesia medication handling needs a new vision. Anesth Analg 2018;126(1):346-350. [doi: 10.1213/ane.0000000000002521]
4. Samost-Williams A, Nanji KC. A systems theoretic process analysis of the medication use process in the operating room. Anesthesiology 2020 Aug;133(2):332-341 [FREE Full text] [doi: 10.1097/ALN.0000000000003376] [Medline: 32541549]
5. Feinglass J, Wang JA, Ye J, Tessier R, Kim H. Hospital care for opioid use in Illinois, 2016-2019. J Behav Health Serv Res 2021 Oct;48(4):597-609 [FREE Full text] [doi: 10.1007/s11414-020-09748-8] [Medline: 33502670]
6. Ye J. Health information system's responses to COVID-19 pandemic in China: a national cross-sectional study. Appl Clin Inform 2021 Mar 19;12(2):399-406 [FREE Full text] [doi: 10.1055/s-0041-1728770] [Medline: 34010976]
7. Institute of Medicine (US) Committee on Quality of Health Care in America. In: Kohn LT, Corrigan JM, Donaldson MS, editors. To Err is Human: Building a Safer Health System. Washington, DC: National Academies Press; 2000.
8. Cooper R, Fogarty-Mack P, Kroll HR, Barach P. Medication safety in anesthesia: epidemiology, causes, and lessons learned in achieving reliable patient outcomes. Int Anesthesiol Clin 2019;57(3):78-95. [doi: 10.1097/AIA.0000000000000232] [Medline: 31577240]
9. Ye J. The impact of electronic health record-integrated patient-generated health data on clinician burnout. J Am Med Inform Assoc 2021 Apr 23;28(5):1051-1056 [FREE Full text] [doi: 10.1093/jamia/ocab017] [Medline: 33822095]
10. Abeysekera A, Bergman IJ, Kluger MT, Short TG. Drug error in anaesthetic practice: a review of 896 reports from the Australian Incident Monitoring Study database. Anaesthesia 2005 Mar;60(3):220-227 [FREE Full text] [doi: 10.1111/j.1365-2044.2005.04123.x] [Medline: 15710005]
11. Berman A. Reducing medication errors through naming, labeling, and packaging. J Med Syst 2004 Feb;28(1):9-29. [doi: 10.1023/b:joms.0000021518.60670.10] [Medline: 15171066]
12. Institute for Safe Medication Practices. ISMP survey on drug shortages for hospital pharmacy directors or their designees only. ISMP Medication Safety Alert! 2017;22(17):5-6.
13. Prielipp R, Magro M, Morell RC, Brull SJ. The normalization of deviance: do we (un)knowingly accept doing the wrong thing? Anesth Analg 2010 May 01;110(5):1499-1502. [doi: 10.1213/ANE.0b013e3181d5adc5] [Medline: 20418310]
14. Jensen LS, Merry AF, Webster CS, Weller J, Larsson L. Evidence-based strategies for preventing drug administration errors during anaesthesia. Anaesthesia 2004 May;59(5):493-504 [FREE Full text] [doi: 10.1111/j.1365-2044.2004.03670.x] [Medline: 15096243]
15. Leahy IC, Lavoie M, Zurakowski D, Baier AW, Brustowicz RM. Medication errors in a pediatric anesthesia setting: incidence, etiologies, and error reduction strategies. J Clin Anesth 2018 Sep;49:107-111 [FREE Full text] [doi: 10.1016/j.jclinane.2018.05.011] [Medline: 29913393]
16. Prakash S, Mullick P, Kumar A, Pawar M. Safe labeling practices to minimize medication errors in anesthesia: 5 case reports and review of the literature. A A Pract 2018 May 15;10(10):261-264. [doi: 10.1213/XAA.0000000000000680] [Medline: 29757795]

XSL•FO
RenderX

17. Maximous R, Wong J, Chung F, Abrishami A. Interventions to reduce medication errors in anesthesia: a systematic review. Can J Anaesth 2021 Jun 01;68(6):880-893. [doi: 10.1007/s12630-021-01959-7] [Medline: 33709263]
18. Larmené-Beld KHM, Alting EK, Taxis K. A systematic literature review on strategies to avoid look-alike errors of labels. Eur J Clin Pharmacol 2018 Aug 12;74(8):985-993 [FREE Full text] [doi: 10.1007/s00228-018-2471-z] [Medline: 29754215]
19. Ye J, Orji IA, Baldridge AS, Ojo TM, Shedul G, Ugwuneji EN, Hypertension Treatment in Nigeria Program Investigators. Characteristics and patterns of retention in hypertension care in primary care settings from the Hypertension Treatment in Nigeria program. JAMA Netw Open 2022 Sep 01;5(9):e2230025 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.30025] [Medline: 36066896]
20. Jelacic S, Bowdle A, Nair BG, Kusulos D, Bower L, Togashi K. A system for anesthesia drug administration using barcode technology: the Codonics Safe Label System and smart anesthesia manager. Anesth Analg 2015 Aug;121(2):410-421. [doi: 10.1213/ANE.0000000000000256] [Medline: 24859078]
21. Ye J, Ren Z. Examining the impact of sex differences and the COVID-19 pandemic on health and health care: findings from a national cross-sectional study. JAMIA Open 2022 Oct;5(3):ooac076 [FREE Full text] [doi: 10.1093/jamiaopen/ooac076] [Medline: 36177395]
22. Merry AF, Webster CS, Hannam J, Mitchell SJ, Henderson R, Reid P, et al. Multimodal system designed to reduce errors in recording and administration of drugs in anaesthesia: prospective randomised clinical evaluation. BMJ 2011 Sep 22;343(sep22 1):d5543-d5543. [doi: 10.1136/bmj.d5543]
23. Martin LD, Grigg EB, Verma S, Latham GJ, Rampersad SE, Martin LD. Outcomes of a failure mode and effects analysis for medication errors in pediatric anesthesia. Paediatr Anaesth 2017 Jun 28;27(6):571-580. [doi: 10.1111/pan.13136] [Medline: 28370645]
24. Ye J, Ma Q. The effects and patterns among mobile health, social determinants, and physical activity: a nationally representative cross-sectional study. AMIA Jt Summits Transl Sci Proc 2021;2021:653-662 [FREE Full text] [Medline: 34457181]
25. Greenwood DA, Litchman ML, Isaacs D, Blanchette JE, Dickinson JK, Hughes A, et al. A new taxonomy for technology-enabled diabetes self-management interventions: results of an umbrella review. J Diabetes Sci Technol 2022 Jul 11;16(4):812-824 [FREE Full text] [doi: 10.1177/19322968211036430] [Medline: 34378424]
26. Ye J, Wang Z, Hai J. Social networking service, patient-generated health data, and population health informatics: national cross-sectional study of patterns and implications of leveraging digital technologies to support mental health and well-being. J Med Internet Res 2022 Apr 29;24(4):e30898 [FREE Full text] [doi: 10.2196/30898] [Medline: 35486428]
27. Ye J, Sanchez-Pinto LN. Three data-driven phenotypes of multiple organ dysfunction syndrome preserved from early childhood to middle adulthood. AMIA Annu Symp Proc 2020;2020:1345-1353 [FREE Full text] [Medline: 33936511]
28. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019 Jan 7;25(1):44-56. [doi: 10.1038/s41591-018-0300-7] [Medline: 30617339]
29. Young IJB, Luz S, Lone N. A systematic review of natural language processing for classification tasks in the field of incident reporting and adverse event analysis. Int J Med Inform 2019 Dec;132:103971. [doi: 10.1016/j.ijmedinf.2019.103971] [Medline: 31630063]
30. Li Q, Fu R, Zhang J, Wang R, Ye J, Xue N, et al. Label-free method using a weighted-phase algorithm to quantitate nanoscale interactions between molecules on DNA microarrays. Anal Chem 2017 Mar 21;89(6):3501-3507. [doi: 10.1021/acs.analchem.6b04596] [Medline: 28230978]
31. Corny J, Rajkumar A, Martin O, Dode X, Lajonchère JP, Billuart O, et al. A machine learning-based clinical decision support system to identify prescriptions with a high risk of medication error. J Am Med Inform Assoc 2020 Nov 01;27(11):1688-1694 [FREE Full text] [doi: 10.1093/jamia/ocaa154] [Medline: 32984901]
32. Koppel R, Metlay JP, Cohen A, Abaluck B, Localio AR, Kimmel SE, et al. Role of computerized physician order entry systems in facilitating medication errors. JAMA 2005 Mar 09;293(10):1197-1203. [doi: 10.1001/jama.293.10.1197] [Medline: 15755942]
33. Rodziewicz T, Houseman B, Hipskind J. Medical Error Reduction and Prevention. Treasure Island, FL: StatPearls Publishing; 2023.
34. Sanuade OA, Ale BM, Baldridge AS, Orji IA, Shedul GL, Ojo TM, Hypertension Treatment in Nigeria Program Investigators. Fixed-dose combination therapy-based protocol compared with free pill combination protocol: results of a cluster randomized trial. J Clin Hypertens (Greenwich) 2023 Feb 20;25(2):127-136 [FREE Full text] [doi: 10.1111/jch.14632] [Medline: 36660886]
35. Eggins S, Slade D. Communication in clinical handover: improving the safety and quality of the patient experience. J Public Health Res 2015 Nov 17;4(3):666 [FREE Full text] [doi: 10.4081/jphr.2015.666] [Medline: 26753165]
36. Ye J. Advancing mental health and psychological support for health care workers using digital technologies and platforms. JMIR Form Res 2021 Jun 30;5(6):e22075 [FREE Full text] [doi: 10.2196/22075] [Medline: 34106874]
37. Yang Y, Rivera AJ, Fortier CR, Abernathy JH. A human factors engineering study of the medication delivery process during an anesthetic: self-filled syringes versus prefilled syringes. Anesthesiology 2016 Apr;124(4):795-803 [FREE Full text] [doi: 10.1097/ALN.0000000000001040] [Medline: 26845139]
38. Berner E. Clinical Decision Support Systems: Theory and Practice. New York, NY: Springer; 2007.

XSL•FO
RenderX

39. Househ M, Ahmad A, Alshaikh A, Alsuweed F. Patient safety perspectives: the impact of CPOE on nursing workflow. Stud Health Technol Inform 2013;183:367-371. [Medline: 23388315]
40. Wulff K, Cummings GG, Marck P, Yurtseven O. Medication administration technologies and patient safety: a mixed-method systematic review. J Adv Nurs 2011 Oct;67(10):2080-2095. [doi: 10.1111/j.1365-2648.2011.05676.x] [Medline: 21545642]
41. Ye J. The role of health technology and informatics in a global public health emergency: practices and implications from the COVID-19 pandemic. JMIR Med Inform 2020 Jul 14;8(7):e19866 [FREE Full text] [doi: 10.2196/19866] [Medline: 32568725]
42. Sakumoto M, Ye J, Kalu R, Jackson KL, Rittner S, Long T, et al. Patient portal perceptions in an urban community health center setting: insights for telehealth. Telehealth Med Today 2022 Nov 23;7(5). [doi: 10.30953/thmt.v7.373]

Abbreviations

AI: artificial intelligence
BCMA: barcode medication administration
CPOE: computerized physician order entry
EHR: electronic health record
eMAR: electronic medication administration record
HFE: human factors engineering
IV: intravenous

Edited by T Leung, J Pearson; submitted 23.10.21; peer-reviewed by A Conway, M Hellaby; comments to author 20.12.21; revised version received 13.03.23; accepted 07.04.23; published 31.05.23.

Please cite as:
Ye J
Patient Safety of Perioperative Medication Through the Lens of Digital Health and Artificial Intelligence
JMIR Perioper Med 2023;6:e34453
URL: https://periop.jmir.org/2023/1/e34453
doi:10.2196/34453
PMID:37256663

©Jiancheng Ye. Originally published in JMIR Perioperative Medicine (http://periop.jmir.org), 31.05.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Perioperative Medicine, is properly cited. The complete bibliographic information, a link to the original publication on http://periop.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

The Role of Wearable Technology in Measuring and Supporting Patient Outcomes Following Total Joint Replacement: Review of the Literature

Gregory Iovanel[1], BSc; David Ayers[1], MD; Hua Zheng[1], PhD

Department of Orthopedics and Physical Rehabilitation, UMass Chan Medical School, Worcester, MA, United States

Corresponding Author:
Hua Zheng, PhD
Department of Orthopedics and Physical Rehabilitation
UMass Chan Medical School
55 Lake Ave N
Worcester, MA, 01655
United States
Phone: 1 774 455 4473
Email: Hua.Zheng@umassmed.edu

Abstract

Background: The incidence rate of total joint replacement (TJR) continues to increase due to the aging population and the surgery that is very successful in providing pain relief to and improving function among patients with advanced knee or hip arthritis. Improving patient outcomes and patient satisfaction after TJR remain important goals. Wearable technologies provide a novel way to capture patient function and activity data and supplement clinical measures and patient-reported outcome measures in order to better understand patient outcomes after TJR.

Objective: We examined the current literature to evaluate the potential role of wearable devices and compare them with existing methods for monitoring and improving patient rehabilitation and outcomes following TJR.

Methods: We performed a literature search by using the research databases supported by the University of Massachusetts Chan Medical School's Lamar Soutter Library, including PubMed and Scopus, supplemented with the Google Scholar search engine. A specific search strategy was used to identify articles discussing the use of wearable devices in measuring and affecting postoperative outcomes of patients who have undergone TJR. Selected papers were organized into a spreadsheet and categorized for our qualitative literature review to assess how wearable data correlated with clinical measures and patient-reported outcome measures.

Results: A total of 9 papers were selected. The literature showed the impact of wearable devices on evaluating and improving postoperative functional outcomes. Wearable-collected data could be used to predict postoperative clinical measures, such as range of motion and Timed Up and Go times. When predicting patient-reported outcomes, specifically Hip Disability and Osteoarthritis Outcome Scores/Knee Injury and Osteoarthritis Outcome Scores and Veterans RAND 12-Item Health Survey scores, strong associations were found between changes in sensor-collected data and changes in patient-reported outcomes over time. Further, the step counts of patients who received feedback from a wearable improved over time when compared to those of patients who did not receive feedback.

Conclusions: These findings suggest that wearable technology has the potential to remotely measure and improve postoperative orthopedic patient outcomes. We anticipate that this review will facilitate further investigation into whether wearable devices are viable tools for guiding the clinical management of TJR rehabilitation.

(JMIR Perioper Med 2023;6:e39396) doi:10.2196/39396

KEYWORDS

total joint replacement; wearables; osteoarthritis; rehabilitation; mobility

XSL•FO
RenderX

Introduction

Total joint replacement (TJR) has proven to be highly effective in relieving joint pain and improving physical function for millions of patients with advanced knee or hip osteoarthritis and continues to be one of the most commonly performed surgical procedures in the United States [1-3]. As this trend persists, increased attention must be paid toward effectively monitoring and coaching patients following surgery to ensure successful rehabilitation. Traditional assessments of postoperative recovery, such as the Timed Up and Go (TUG) and 6-minute walk tests, are considered gold standards for measuring mobility, balance, and walking ability [4]. However, these assessments require in-person monitoring by health care providers and do not replicate activities of daily living. Patient-reported outcomes (PROs) have been widely used to evaluate joint pain and physical function through standardized patient questionnaires. Patients report on how they perceive their health status without the interpretation of a medical professional. Although the assessment of PROs has become part of the standard of care in many orthopedic practices, the implementation of PRO capture, the maintenance of data integrity, data interpretation, and cost management are still challenging for many practices [5-9]. The internet-based remote monitoring of patient mobility data is an alternative method of collecting patient data following surgery that has recently been introduced and warrants further evaluation.

Wearable technologies provide a novel way to capture patient function and activity data and supplement clinical measures and PRO measures (PROMs) to better understand patient recovery after TJR. Wearable technologies, in the context of health care, refer to devices that can record real-time data from an individual while worn. These devices include accelerometers, which capture the acceleration of a limb or the entire body; gyroscopes, which measure orientation and angular velocity; and inertial measurement units—a more sophisticated technology that combines an accelerometer, gyroscope, and magnetometer and is capable of reporting the movement, orientation, and position in space of a person or object [10]. Many companies manufacture such devices that can be synced to a smartphone, computer, or tablet to transmit patient mobility data securely and instantly to health care providers via an internet-based application. Medical professionals are then able to track patients' progress in real time and tailor rehabilitation regimens for patients to follow, based on the data obtained [11,12]. Such wearable technologies could offer the possibility of capturing real-time function data on the rehabilitation and recovery of patients who have undergone TJR and eliminating the need for direct supervision. In addition, a connected mobile app can be developed to collect PROMs, thereby minimizing the need for additional PROM capture tools [13]. Current research has shown the feasibility of wearable devices and their capability for motion and activity tracking [14]. However, it is not clear whether the activity data collected by wearable devices can serve as outcome measures or as adjuncts to support outcome monitoring. There is a dearth of consensus on whether wearables can be used as effective tools, can be aligned with standard clinical measures and PROMs, or can even improve outcomes.

To promote wearable use as part of rehabilitation programs following TJR, their impact on postoperative patient outcomes, as well as their accuracy in measuring these outcomes, must be further investigated. This paper seeks to review the current landscape of orthopedic wearables literature and assess the effectiveness of available devices with respect to evaluating and improving postoperative outcomes.

Methods

A literature search was conducted by using the research databases supported by the University of Massachusetts Chan Medical School's Lamar Soutter Library, including PubMed and Scopus, supplemented with the Google Scholar search engine. Articles published in English from 2004 to 2021 were reviewed. The search terms used to identify these articles are defined in Textbox 1.

Textbox 1. Literature search strategy (search terms used in the literature search strategy).

Term groupings and search terms

- Wearable devices
  - *("wearable") AND ("devices" OR "technology")*
- Patient
  - *("total joint replacement" OR "total knee replacement" OR "total hip replacement") AND "outcomes"*
- Rehabilitation
  - *"rehabilitation" OR "recovery"*

The inclusion criteria included English-language articles, research studies, and studies with wearable technology that focused on comparing wearable-collected data with clinical measures or PROMs or affecting patient outcomes. The exclusion criteria were articles focusing on wearable device design, study protocols, theoretical articles, books, or book chapters. Titles and abstracts of identified articles were screened to determine eligibility based on the inclusion and exclusion criteria. Since only a limited number of papers met the inclusion criteria, a full reading was conducted for all of the eligible papers.

The information was tabulated via a standardized Excel (Microsoft Corporation) form that was developed for this review,

XSL•FO
RenderX

which included the first author's name, year of publication, name and type of the wearable device, location where the device was worn, number of patients in the study, outcome measures, and study findings (Table 1). A narrative literature review of the selected articles was conducted by 2 reviewers, providing a qualitative overview of outcome measures, data collection methods, and main findings.

XSL•FO
RenderX

Table 1. Classification of selected articles (papers were organized by the first author's name, year, wearable device, device type, device location, number of patients, outcome measures, and findings).

Authors, year	Wearable device	Device type	Device location	Patients, n	Outcome measure	Findings
Kwasnicki et al [15], 2015	e-AR (Imperial College London)	Accelerometer	Ear	14	TUG[a] time and ROM[b]	The classification of patients into preoperative, normal, and 24-week postoperative groups based on outcomes was 89% accurate, while classification for all time intervals was 69% accurate.
Chiang et al [16], 2017	APDM OPAL (APDM Wearable Technologies)	Accelerometer, gyroscope, magnetometer, and barometer	Thigh and calf (2 sensors)	18	Satisfaction	Only 17% of patients felt uncomfortable with the sensor belt.
Bendich et al [17], 2019	Fitbit Flex (Fitbit LLC)	Accelerometer	Wrist	22	Daily step count, daily minutes active, HOOS/KOOS[c], and VR-12[d] score	Changes from preoperative levels to 6-week postoperative levels in "daily step count" and "daily minutes active" (collected with a wearable sensor) were strongly associated with improvements in HOOSs/KOOSs and VR-12 physical component scores (collected over the same period).
Chen et al [18], 2015	APDM OPAL	Accelerometer, gyroscope, and magnetometer	Chest, thigh, and calf (3 sensors)	10	ROM	The device was able to identify proper exercise posture 88.26% of the time.
Battenberg et al [19], 2017	Fitbit One (Fitbit LLC), Omron HJ-321 (Omron Corporation), Sportline 340 Strider (Sportline Inc), Fitbit Force (Fitbit LLC), Nike+ Fuelband SE (Nike Inc), and StepWatch Activity Monitor (Orthocare Innovations)	Fitbit One (accelerometer), Omron HJ-321 (pedometer and accelerometer), Sportline 340 Strider (pedometer), Fitbit Force (accelerometer), Nike+ Fuelband SE (accelerometer), and StepWatch Activity Monitor (accelerometer)	Fitbit One (waist), Omron HJ-321 (waist), Sportline 340 Strider (waist), Fitbit Force (wrist), Nike+ Fuelband SE (wrist), and StepWatch Activity Monitor (ankle)	30	Step count	The waist-based devices—Fitbit One and Omron HJ-321—were >90% accurate in counting steps for all activities, the wristband devices were <90% accurate for most activities, and the StepWatch Activity Monitor (ankle) was >95% accurate for lower cadence activities but undercounted running by 25%.
Toogood et al [20], 2016	Fitbit (Fitbit LLC)	Accelerometer	Ankle	33	Compliance	The mean compliance over 30 days was 26.7 days (89%).
Saporito et al [21], 2019	Custom	Accelerometer and barometer	Neck (pendant)	15	TUG time	A strong correlation (ρ=0.70) was observed between remote TUG times and standardized TUG times.
Van der Walt et al [22], 2018	Garmin Vivofit 2 (Garmin Ltd)	Accelerometer	Wrist	163	Step count	Participants receiving feedback on step goals from the device had significantly higher (P<.03) mean daily step counts than those of participants who did not receive any feedback from the device.
Kuiken et al [23], 2004	Custom	Goniometer	Knee	11	ROM and mean activity rate	After total knee arthroplasty, patients wearing a device providing feedback had higher mean total activity rates—a measure of ROM—on days when they did not receive feedback from the device (mean 22.5, SD 11.1 activity counts per hour) than on days when they did receive feedback (mean 15.1, SD 10.9 activity counts per hour), but this was not statistically significant (P=.11).

XSL•FO
RenderX

[a]TUG: Timed Up and Go.

[b]ROM: range of motion.

[c]HOOS/KOOS: Hip Disability and Osteoarthritis Outcome Score/Knee Injury and Osteoarthritis Outcome Score.

[d]VR-12: Veterans RAND 12-Item Health Survey.

The standard postoperative TJR outcome measures in this literature review included (1) assessments typically conducted in clinical settings, such as range of motion (ROM) assessments and the TUG test, and (2) PROMs, such as joint-specific outcome measures (Hip Disability and Osteoarthritis Outcome Score/Knee Injury and Osteoarthritis Outcome Score [HOOS/KOOS]), global health measures (Veterans RAND 12-Item Health Survey [VR-12]), patient satisfaction, and activity adherence.

Results

A total of 9 articles that met the inclusion criteria were identified. The articles evaluated the mobility and activity data collected through the wearable devices and compared them with standard clinical outcome measures and PROMs.

Correlation of Wearables and Clinical Measures

In evaluating ROM and TUG time, the wearables varied in accuracy. Kwasnicki et al [15] observed 14 patients who underwent total knee replacement and wore the e-AR accelerometer (Imperial College London) on the ear to conduct home-based mobility assessments. The authors compared a generated sensor score, which was based on sensor data, with the results of other assessment techniques (TUG test and knee ROM). They calculated Spearman ρ correlation coefficients between sensor scores and TUG and ROM measurements to assess the strength of association. They found that perioperative sensor scores correlated, albeit not significantly for all activities, with TUG time and ROM improvements. In another study that focused on TUG measurements, Saporito et al [21] collected standardized TUG data from 239 community-living older adults in a laboratory and sensor-based data on participants' activities of daily living through a wearable pendant device for at least 3 days and developed a regularized linear model for estimating remote TUG times. Based on the device data of 15 patients who underwent total hip replacement, a strong correlation was observed between estimated remote TUG times and standardized TUG times via leave-one-out cross-validation.

Correlation of Wearables and PROMs

Data from wearable devices may correlate with PROMs. Bendich et al [17] aimed to determine whether sensor-collected data could be used as predictors of PROMs. In their study, 22 patients who underwent TJR wore a Fitbit Flex (Fitbit LLC) device on the wrist, which allowed for the observation of potential associations between "daily step count" and "daily minutes active" data collected by the wearable and PROMs, specifically the HOOS/KOOS and VR-12, over time. The researchers found that changes observed in "daily step count" from before the operation to postoperative week 6 were strongly associated with changes in VR-12 scores, while changes observed in "daily minutes active" from before the operation to postoperative week 6 were strongly associated with changes in HOOSs/KOOSs.

Impact of the Use of Wearables on Patient Outcomes

The authors of 2 articles discussed the impact of the use of wearable devices on postoperative TJR patient outcomes. Specifically, the researchers investigated how the ability of devices to offer feedback on exercise and rehabilitation to patients may impact patient outcomes. Van der Walt et al [22] randomized 163 patients who underwent TJR into 2 groups; one received feedback for their rehabilitation via the Garmin Vivofit 2 (Garmin Ltd) accelerometer, and the other did not receive any feedback. They found that the mean daily step counts of the group that received feedback were significantly higher than those of the group that did not receive feedback (43% higher in postoperative week 1, 33% higher in postoperative week 2, 21% higher in postoperative week 6, and 17% higher at postoperative month 6). Surprisingly, in a study with 11 patients who underwent total knee arthroplasty, Kuiken et al [23] found that patients who wore a device that provided feedback had a slightly higher mean total activity rate on days when they did not receive feedback from the device compared to that on days when they did receive feedback from the device, although this difference was not statistically significant.

Patients reported high satisfaction with and adherence for the use of wearable devices. A study by Chiang et al [16] found that in a group of 18 patients who underwent total knee replacement and wore a thigh- and calf-worn wearable, 83% reported no discomfort when wearing the device. In a study by Toogood et al [20] on device adherence, the mean compliance rate for wearing an ankle-based Fitbit accelerometer (Fitbit LLC) among 33 patients who underwent total hip replacement was 89% (26.7/30 days). Although this study noted that devices were worn for 24 hours per day, apart from during washing, the daily duration of use was not specifically mentioned in the other selected studies.

Device Data Accuracy Evaluation

Several devices were found to be generally accurate in counting steps. Battenberg et al [19] tested the accuracy of several widely used wearable devices in a convenience sample of 30 healthy participants. They found that the waist-worn Fitbit One (Fitbit LLC) and Omron HJ-321 (Omron Corporation) had greater than 90% accuracy in step counting during all activities; the wristband devices, such as the Fitbit Force (Fitbit LLC) and Nike+ Fuelband SE (Nike Inc), had less than 90% accuracy for most activities; and the ankle-worn StepWatch Activity Monitor (Orthocare Innovations) was greater than 95% accurate when counting steps during lower cadence activities but undercounted steps during running by 25%. In a study by Chen at al [18], 10 healthy participants, while wearing 3 APDM OPAL (APDM Wearable Technologies) sensors on the chest, thigh, and calf, performed 3 different rehabilitation exercises that were designed for patients with knee osteoarthritis to manage rehabilitation progress at home. The device was found to have an overall recognition accuracy of 97% for exercise type classification

XSL•FO
RenderX

and an overall recognition accuracy of 88% for proper exercise posture.

Discussion

Wearable Data Can Be Used as Alternative Outcome Measures

Postoperative TJR recovery remains a black box to health care providers until patients report to a clinic or respond to a survey. With adequate implementation and the ability to collect data continuously, even from a remote setting, wearable devices can help health care providers to monitor progress consistently and detect early problems in rehabilitation [24]. The literature shows that function and activity data obtained from wearables, including step count and exercise tracking data, correlate with both clinical outcomes and PROMs [15,17,21]. Such wearable data are able to provide measures of patients' objective functional outcomes that are comparable with standard clinical metrics and patient surveys. In addition, the opportunity to regularly monitor patients in real time and allow for direct feedback from and communication with health care providers can alleviate the inconveniences of unnecessary office visits and costs; patients with good progress can continue at-home rehabilitation, while patients with poor progress can be alerted to proactively visit a clinic before permanent complications occur. Further research is however needed to evaluate device bias and data accuracy to make sure that wearable results are reliable.

There has also been some support in the literature for the use of monitoring insoles, particularly for the purpose of load and gait analysis. Although preliminary findings suggest that monitoring insoles have good accuracy in measuring foot load distribution and natural gait, the few studies that have been performed are limited by small sample sizes [25,26]. Additional investigations with larger data sets will be needed.

Wearables Can Be Used to Improve Outcomes

In addition to generating data that correlate with established outcomes, wearables can also be used to improve outcomes overall by more actively engaging patients in exercise and activity [24,27]. Indeed, devices connected to mobile apps can provide feedback to patients regarding their rehabilitation routines, and the mobility metrics, such as daily step count, of patients who received such feedback significantly improved when compared to those of patients who did not receive feedback [22]. Additionally, the ability of these wearables to provide daily exercise reminders to patients and plot their progress over time sustained patients' motivation and further contributed to outcome improvement [28,29].

Wearables and Apps Can Be Included in Future Health IT Infrastructure

As orthopedic clinical research has progressed, more data sources have emerged from which to monitor and guide patient rehabilitation and care following TJR. Whereas most patient data previously originated from electronic health records, direct patient-generated data in the form of PROMs or outcomes tracked and collected by wearables aptly supplement clinically collected data. Particularly, the ability of wearables to generate objective, continuous data showing trends in patient progress is unique in comparison to PROMs, which provide subjective data from predetermined time points, and electronic health record data, which are only collected during patients' point-of-care visits and require medical professionals' involvement. Moreover, with the increased emphasis on telemedicine, particularly since the COVID-19 pandemic, the remote monitoring of patient recovery via wearables represents a potential new path toward collecting patient data and guiding clinical decision-making [30,31]. These novel applications emphasize the role of wearables in the future of health IT infrastructure.

Challenges

There are still challenges to the implementation of wearable technology. Technical support will be needed for device calibration and data collection. Some research teams have assisted in the use of wearables during appointments scheduled at patients' homes [15], hospital wards, or outpatient clinics [16]. Patients also need to be provided with training and guidance before and during the study period to ensure proper device mounting and use. Additionally, standardization must be established across different devices and across data collection in different settings to ensure that data are comparable and meaningful.

Conclusion

This review discusses the current state of the literature regarding the effectiveness of wearable devices in measuring and improving TJR outcomes, as well as the future directions of wearable device use. Wearable technologies have great potential for assessing and enhancing patients' postoperative physical function. Wearables can be effective, alternative tools for evaluating TJR outcomes, as early findings have shown correlations among wearable-recorded data, PROMs, and clinical outcomes. The implementation and standardization of wearables should be addressed in future research.

Acknowledgments

The authors thank Sylvie Puig, PhD, for her editorial assistance.

Conflicts of Interest

DA is a scientific advisor to Exactech.

References

XSL•FO
RenderX

1. Williams SN, Wolford ML, Bercovitz A. Hospitalization for total knee replacement among inpatients aged 45 and over: United States, 2000-2010. NCHS Data Brief 2015 Aug(210):1-8 [FREE Full text] [Medline: 26375255]
2. Kurtz S, Ong K, Lau E, Mowat F, Halpern M. Projections of primary and revision hip and knee arthroplasty in the United States from 2005 to 2030. J Bone Joint Surg Am 2007 Apr;89(4):780-785. [doi: 10.2106/JBJS.F.00222] [Medline: 17403800]
3. Zagra L. Advances in hip arthroplasty surgery: what is justified? EFORT Open Rev 2017 May 11;2(5):171-178 [FREE Full text] [doi: 10.1302/2058-5241.2.170008] [Medline: 28630755]
4. Ko V, Naylor JM, Harris IA, Crosbie J, Yeo AET. The six-minute walk test is an excellent predictor of functional ambulation after total knee arthroplasty. BMC Musculoskelet Disord 2013 Apr 24;14:145 [FREE Full text] [doi: 10.1186/1471-2474-14-145] [Medline: 23617377]
5. Collins NJ, Roos EM. Patient-reported outcomes for total hip and knee arthroplasty: commonly used instruments and attributes of a "good" measure. Clin Geriatr Med 2012 Aug;28(3):367-394. [doi: 10.1016/j.cger.2012.05.007] [Medline: 22840304]
6. Franklin PD, Lewallen D, Bozic K, Hallstrom B, Jiranek W, Ayers DC. Implementation of patient-reported outcome measures in U.S. Total joint replacement registries: rationale, status, and plans. J Bone Joint Surg Am 2014 Dec 17;96(Suppl 1):104-109 [FREE Full text] [doi: 10.2106/JBJS.N.00328] [Medline: 25520425]
7. Ayers DC. Implementation of patient-reported outcome measures in total knee arthroplasty. J Am Acad Orthop Surg 2017 Feb;25 Suppl 1:S48-S50. [doi: 10.5435/JAAOS-D-16-00631] [Medline: 27941415]
8. Lavallee DC, Chenok KE, Love RM, Petersen C, Holve E, Segal CD, et al. Incorporating patient-reported outcomes into health care to engage patients and enhance care. Health Aff (Millwood) 2016 Apr;35(4):575-582. [doi: 10.1377/hlthaff.2015.1362] [Medline: 27044954]
9. Porter I, Gonçalves-Bradley D, Ricci-Cabello I, Gibbons C, Gangannagaripalli J, Fitzpatrick R, et al. Framework and guidance for implementing patient-reported outcomes in clinical practice: evidence, challenges and opportunities. J Comp Eff Res 2016 Aug;5(5):507-519 [FREE Full text] [doi: 10.2217/cer-2015-0014] [Medline: 27427277]
10. Zügner R, Tranberg R, Timperley J, Hodgins D, Mohaddes M, Kärrholm J. Validation of inertial measurement units with optical tracking system in patients operated with total hip arthroplasty. BMC Musculoskelet Disord 2019 Feb 06;20(1):52 [FREE Full text] [doi: 10.1186/s12891-019-2416-4] [Medline: 30727979]
11. Ianculescu M, Andrei B, Alexandru A. A smart assistance solution for remotely monitoring the orthopaedic rehabilitation process using wearable technology: re.flex system. Studies in Informatics and Control 2019;28(3):317-326 [FREE Full text] [doi: 10.24846/v28i3y201908]
12. Smuck M, Odonkor CA, Wilt JK, Schmidt N, Swiernik MA. The emerging clinical role of wearables: factors for successful implementation in healthcare. NPJ Digit Med 2021 Mar 10;4(1):45 [FREE Full text] [doi: 10.1038/s41746-021-00418-3] [Medline: 33692479]
13. Liao Y, Thompson C, Peterson S, Mandrola J, Beg MS. The future of wearable technologies and remote monitoring in health care. Am Soc Clin Oncol Educ Book 2019 Jan;39:115-121 [FREE Full text] [doi: 10.1200/EDBK_238919] [Medline: 31099626]
14. Argent R, Slevin P, Bevilacqua A, Neligan M, Daly A, Caulfield B. Wearable sensor-based exercise biofeedback for orthopaedic rehabilitation: A mixed methods user evaluation of a prototype system. Sensors (Basel) 2019 Jan 21;19(2):432 [FREE Full text] [doi: 10.3390/s19020432] [Medline: 30669657]
15. Kwasnicki RM, Ali R, Jordan SJ, Atallah L, Leong JJH, Jones GG, et al. A wearable mobility assessment device for total knee replacement: A longitudinal feasibility study. Int J Surg 2015 Jun;18:14-20 [FREE Full text] [doi: 10.1016/j.ijsu.2015.04.032] [Medline: 25868424]
16. Chiang CY, Chen KH, Liu KC, Hsu SJP, Chan CT. Data collection and analysis using wearable sensors for monitoring knee range of motion after total knee arthroplasty. Sensors (Basel) 2017 Feb 22;17(2):418 [FREE Full text] [doi: 10.3390/s17020418] [Medline: 28241434]
17. Bendich I, Chung C, Hwang K, Patterson J, Mulvihill J, Barry J, et al. Changes in prospectively collected longitudinal patient-generated health data are associated with short-term patient-reported outcomes after total joint arthroplasty: a pilot study. Arthroplast Today 2019 Mar 14;5(1):61-63 [FREE Full text] [doi: 10.1016/j.artd.2019.01.005] [Medline: 31020024]
18. Chen KH, Chen PC, Liu KC, Chan CT. Wearable sensor-based rehabilitation exercise assessment for knee osteoarthritis. Sensors (Basel) 2015 Feb 12;15(2):4193-4211. [doi: 10.3390/s150204193] [Medline: 25686308]
19. Battenberg AK, Donohoe S, Robertson N, Schmalzried TP. The accuracy of personal activity monitoring devices. Semin Arthroplasty 2017 Jun;28(2):71-75. [doi: 10.1053/j.sart.2017.07.006]
20. Toogood PA, Abdel MP, Spear JA, Cook SM, Cook DJ, Taunton MJ. The monitoring of activity at home after total hip arthroplasty. Bone Joint J 2016 Nov;98-B(11):1450-1454. [doi: 10.1302/0301-620X.98B11.BJJ-2016-0194.R1] [Medline: 27803219]
21. Saporito S, Brodie MA, Delbaere K, Hoogland J, Nijboer H, Rispens SM, et al. Remote timed up and go evaluation from activities of daily living reveals changing mobility after surgery. Physiol Meas 2019 Apr 03;40(3):035004. [doi: 10.1088/1361-6579/ab0d3e] [Medline: 30840937]

XSL•FO
RenderX

22. Van der Walt N, Salmon LJ, Gooden B, Lyons MC, O'Sullivan M, Martina K, et al. Feedback from activity trackers improves daily step count after knee and hip arthroplasty: A randomized controlled trial. J Arthroplasty 2018 Nov;33(11):3422-3428. [doi: 10.1016/j.arth.2018.06.024] [Medline: 30017217]
23. Kuiken TA, Amir H, Scheidt RA. Computerized biofeedback knee goniometer: acceptance and effect on exercise behavior in post-total knee arthroplasty rehabilitation. Arch Phys Med Rehabil 2004 Jun;85(6):1026-1030. [doi: 10.1016/j.apmr.2003.08.088] [Medline: 15179662]
24. Farr-Wharton G, Li J, Hussain M, Freyne J. Mobile supported health services: Experiences in orthopaedic care. 2020 Jul 28 Presented at: 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS); July 28-30, 2020; Rochester, Minnesota. [doi: 10.1109/cbms49503.2020.00074]
25. Chatzaki C, Skaramagkas V, Tachos N, Christodoulakis G, Maniadi E, Kefalopoulou Z, et al. The smart-insole dataset: Gait analysis using wearable sensors with a focus on elderly and Parkinson's patients. Sensors (Basel) 2021 Apr 16;21(8):2821 [FREE Full text] [doi: 10.3390/s21082821] [Medline: 33923809]
26. Cavalleri M, Reni G. Active monitoring insole: A wearable device for monitoring foot load distribution in home-care context. 2008 Presented at: 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; August 20-25, 2008; Vancouver, British Columbia. [doi: 10.1109/iembs.2008.4650198]
27. Bahadori S, Immins T, Wainwright TW. A review of wearable motion tracking systems used in rehabilitation following hip and knee replacement. J Rehabil Assist Technol Eng 2018 Jun 18;5:2055668318771816 [FREE Full text] [doi: 10.1177/2055668318771816] [Medline: 31191937]
28. Chen HC, Chuang TY, Lin PC, Lin YK, Chuang YH. Effects of messages delivered by mobile phone on increasing compliance with shoulder exercises among patients with a frozen shoulder. J Nurs Scholarsh 2017 Jul;49(4):429-437. [doi: 10.1111/jnu.12308] [Medline: 28632975]
29. Chen YP, Lin CY, Tsai MJ, Chuang TY, Lee OKS. Wearable motion sensor device to facilitate rehabilitation in patients with shoulder adhesive capsulitis: Pilot study to assess feasibility. J Med Internet Res 2020 Jul 23;22(7):e17032 [FREE Full text] [doi: 10.2196/17032] [Medline: 32457026]
30. Bini SA, Schilling PL, Patel SP, Kalore NV, Ast MP, Maratt JD, et al. Digital orthopaedics: A glimpse into the future in the midst of a pandemic. J Arthroplasty 2020 Jul;35(7S):S68-S73 [FREE Full text] [doi: 10.1016/j.arth.2020.04.048] [Medline: 32416956]
31. Lim MA, Pranata R. Teleorthopedic: A promising option during and after the coronavirus disease 2019 (COVID-19) pandemic. Front Surg 2020 Aug 28;7:62 [FREE Full text] [doi: 10.3389/fsurg.2020.00062] [Medline: 33005624]

Abbreviations

HOOS/KOOS: Hip Disability and Osteoarthritis Outcome Score/Knee Injury and Osteoarthritis Outcome Score
PRO: patient-reported outcome
PROM: patient-reported outcome measure
ROM: range of motion
TJR: total joint replacement
TUG: Timed Up and Go
VR-12: Veterans RAND 12-Item Health Survey

Edited by J Pearson; submitted 09.05.22; peer-reviewed by AA Seid, M Kraus, K Alexander; comments to author 28.06.22; revised version received 14.11.22; accepted 13.12.22; published 12.01.23.

Please cite as:
Iovanel G, Ayers D, Zheng H
The Role of Wearable Technology in Measuring and Supporting Patient Outcomes Following Total Joint Replacement: Review of the Literature
JMIR Perioper Med 2023;6:e39396
URL: https://periop.jmir.org/2023/1/e39396
doi:10.2196/39396
PMID:36633891

©Gregory Iovanel, David Ayers, Hua Zheng. Originally published in JMIR Perioperative Medicine (http://periop.jmir.org), 12.01.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Perioperative Medicine, is properly cited. The complete bibliographic

XSL•FO
RenderX

information, a link to the original publication on http://periop.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Accuracy of Wrist-Worn Photoplethysmogram–Measured Heart and Respiratory Rates in Abdominal Surgery Patients: Observational Prospective Clinical Validation Study

Jonna A van der Stam[1,2,3], MSc; Eveline H J Mestrom[4], MD; Jai Scheerhoorn[5], MD; Fleur E N B Jacobs[6], MSc; Simon Nienhuijs[5], MD, PhD; Arjen-Kars Boer[2,3], PhD; Natal A W van Riel[1,3,7], PhD; Helma M de Morree[8], PhD; Alberto G Bonomi[8], PhD; Volkher Scharnhorst[1,2,3], PhD; R Arthur Bouwman[4,9], MD, PhD

[1]Department of Biomedical Engineering, Eindhoven University of Technology, Eindhoven, Netherlands
[2]Clinical Laboratory, Catharina Hospital, Eindhoven, Netherlands
[3]Expert Center Clinical Chemistry Eindhoven, Eindhoven, Netherlands
[4]Department of Anesthesiology, Intensive Care & Pain Medicine, Catharina Hospital, Eindhoven, Netherlands
[5]Department of Surgery, Catharina Hospital, Eindhoven, Netherlands
[6]Department of Medical Physics, Catharina Hospital, Eindhoven, Netherlands
[7]Department of Vascular Medicine, Amsterdam University Medical Centers, Amsterdam, Netherlands
[8]Patient Care & Monitoring Department, Philips Research, Eindhoven, Netherlands
[9]Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, Netherlands

Corresponding Author:
Jonna A van der Stam, MSc
Department of Biomedical Engineering
Eindhoven University of Technology
de Zaale
Eindhoven, 5612 AZ
Netherlands
Phone: 31 40 2398675
Email: jonna.vd.stam@catharinaziekenhuis.nl

Abstract

Background: Postoperative deterioration is often preceded by abnormal vital parameters. Therefore, vital parameters of postoperative patients are routinely measured by nursing staff. Wrist-worn sensors could potentially provide an alternative tool for the measurement of vital parameters in low-acuity settings. These devices would allow more frequent or even continuous measurements of vital parameters without relying on time-consuming manual measurements, provided their accuracy in this clinical population is established.

Objective: This study aimed to assess the accuracy of heart rate (HR) and respiratory rate (RR) measures obtained via a wearable photoplethysmography (PPG) wristband in a cohort of postoperative patients.

Methods: The accuracy of the wrist-worn PPG sensor was assessed in 62 post–abdominal surgery patients (mean age 55, SD 15 years; median BMI 34, IQR 25-40 kg/m^2). The wearable obtained HR and RR measurements were compared to those of the reference monitor in the postanesthesia or intensive care unit. Bland-Altman and Clarke error grid analyses were performed to determine agreement and clinical accuracy.

Results: Data were collected for a median of 1.2 hours per patient. With a coverage of 94% for HR and 34% for RR, the device was able to provide accurate measurements for the large majority of the measurements as 98% and 93% of the measurements were within 5 bpm or 3 rpm of the reference signal. Additionally, 100% of the HR and 98% of the RR measurements were clinically acceptable on Clarke error grid analysis.

Conclusions: The wrist-worn PPG device is able to provide measurements of HR and RR that can be seen as sufficiently accurate for clinical applications. Considering the coverage, the device was able to continuously monitor HR and report RR when measurements of sufficient quality were obtained.

Trial Registration: ClinicalTrials.gov NCT03923127; https://www.clinicaltrials.gov/ct2/show/NCT03923127

XSL•FO
RenderX

(JMIR Perioper Med 2023;6:e40474) doi:10.2196/40474

KEYWORDS

Telemetry; monitoring; photoplethysmography; PPG; photoplethysmogram; wearable monitoring; vital parameter; wearable sensor; sensor; heart rate; respiratory Rate; respiration; respiratory; breathing; monitoring; wearable; postoperative; post-operative; vital sign

Introduction

Alterations in vital parameters can often be found hours before a life-threatening event occurs [1-7]. In current clinical practice, postoperative monitoring often consists of a period of continuous monitoring in an intensive care or postanesthesia care unit, followed by an admission to a general ward. Since continuous monitoring of vital parameters is not present in the general ward, nursing staff performs the so-called spot checks to monitor the patient's vital parameters. During these spot checks, the nursing staff measures several vital parameters, often followed by manual entry or calculation of an early warning score such as the Modified Early Warning Score, to identify patients at risk of deterioration [8]. In clinical practice, these spot checks form a considerable workload, and vital parameters, especially respiratory rate (RR), are often poorly registered [9,10]. Additionally, as the name implies, these spot checks capture only vital parameters at a specific moment in time, and vital parameters during the rest of the day remain unknown. Alternatively, wearable sensors could be used to unobtrusively and continuously measure vital parameters in postoperative patients. However, their accuracy in postoperative patients should be established prior to introduction in clinical practice.

One type of wearable sensor that can monitor a patient's vital parameters is a photoplethysmography (PPG) wristband. This type of sensor measures the intensity of the light reflected from the skin, which indicates changes in the blood volume in peripheral circulation, to determine both heart rate (HR) and RR [11]. Wrist-worn PPG sensors have potential for use as a continuous, unobtrusive monitoring system in low-acuity settings such as the general ward.

A few studies have reported the accuracy of other PPG-based wearables in hospitalized patients; however, these trials only studied the measurement of HR [12-15]. Additionally, the accuracy of wrist plethysmography devices for HR measurements in a perioperative cohort was previously investigated and found to be clinically acceptable [15]. However, as both HR and RR have been identified as important parameters for the prediction of clinical deterioration, accuracy for both vital parameters should be established [16]. Therefore, this study aims to assess the accuracy of a wrist-worn PPG device for measuring both RR and HR in postoperative patients.

Methods

Study Population

These analyses were performed with a subpopulation of Transitional Care Study 3 (TRICA; ClinicalTrials.gov NCT03923127)—a single-center study on wearable monitoring in postoperative patients in a tertiary hospital [17,18]. All adult patients scheduled for major abdominal oncological or bariatric surgery from April 2019 to August 2020 who were willing and able to sign informed consent were eligible for participation. Patients who met any of the following criteria were not included: being pregnant or breastfeeding, having an allergy to tissue adhesives, having an antibiotic-resistant skin infection, having an active implantable device, or having any skin condition at the area of application of the devices. This subanalysis describes 68 postoperative patients, and inclusion into this subanalysis for accuracy of the wearable sensor was based on the availability of research personnel and real-time data logging equipment.

Ethics Approval

The trial was approved by the medical ethical committee METC Máxima MC, Veldhoven, The Netherlands (W19.001).

Data Collection

The wearable PPG wristband device, ELAN, was equipped with a Philips Cardio and Motion Monitoring Module (CM3, Philips Electronic Nederland BV), which contains a PPG and 3-axial accelerometer sensor. The PPG sensor measures the intensity of the green light scatter-reflected from the skin to determine changes in blood volume in the peripheral circulation with a sampling frequency of 32 Hz [19]. From the obtained PPG signal, HR and RR were determined using previously published algorithms, the RR measurements are derived from interbeat interval variability and PPG amplitude [20]. Additionally, the device reports a quality index with each measured vital value, which mostly captures the signal-to-noise ratio [15]. Only vitals with a quality index of 4 (range 0-4), are considered to be of high quality and can be included in further analysis.

Shortly after surgery, the PPG wristband was applied to the patient's wrist in the postanesthesia care unit (PACU) or intensive care unit (ICU), depending on where the patient was recovering immediately after surgery. The wristband then continuously collected both HR and RR.

As a ground truth, the electrocardiogram (ECG)-based HR and capnography-based RR signals of 68 patients were extracted from the bedside monitor in the PACU or ICU. These signals were saved in real time for offline processing, allowing comparison between the HR and RR measured by the PPG wristband and the reference monitor. In the PACU, vital parameters from the CAR-ESCAPE monitor B650 (GE Healthcare) were extracted using iCollect software (GE Healthcare) with a sampling frequency of 250 Hz for ECG and 1 Hz for RR. In the ICU, vital parameters were extracted from the Philips IntelliVue MP70 monitor using IntelliVue software (Philips) with a sampling frequency of 100 Hz for ECG and 0.1 Hz for RR. HR was derived from the ECG on second-to-second bases using QRS detection algorithms, RR was obtained using the patient monitors' algorithms.

XSL•FO
RenderX

Data Analysis

The obtained vital parameters from the PPG wristband and the reference monitors were synchronized using a means of cross-correlation on the HR signals, and synchronized signals were visually inspected and corrected if necessary. Patients with a reference recording length shorter than 15 minutes were excluded from the analysis.

Low-quality measurements were excluded from both the PPG and monitor data. For the PPG wristband vitals, a low quality index can originate from motion artefacts or a low signal-to-noise ratio. For HR and RR, detection of arrhythmia using an arrhythmia detection algorithm would also lead to a low quality score [21]. For the reference monitor, the logged ECG and capnography signals were visually inspected to identify low-quality measurements, based on assessment of the temporal sequence.

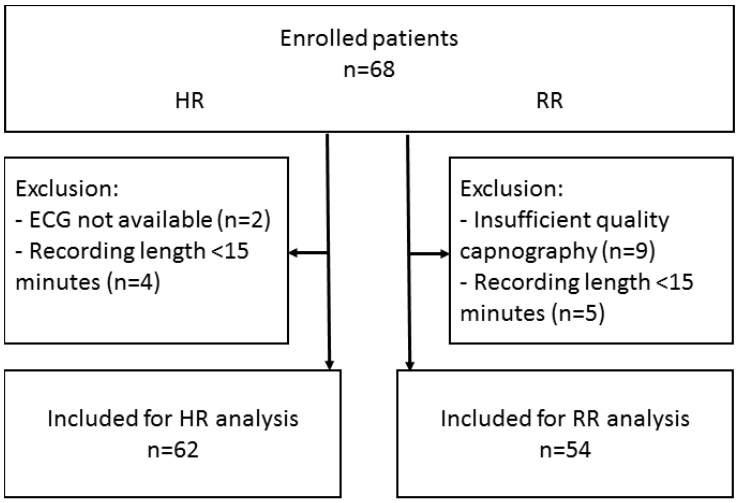
Baseline characteristics are expressed as mean (SD) or, in case of nonnormally distributed values, as median (IQR) values. Agreement between the PPG wristband and reference monitor measurements on a second-to-second basis was visualized using Bland-Altman plots [22]. As multiple observations from the same patients were analyzed, the bias and limits of agreement were calculated using the method for repeated measures of Zou et al [23]. Additionally, the 95% CIs around the limits of agreement were assessed using MOVER [23].

According to the American National Standards Institute consensus standard, the error for HR measurements should be ≤10% or ≤5 bpm. In this analysis, an error of ≤5 bpm for HR and ≤3 rpm for RR was considered clinically acceptable. Additionally, Clarke error grid analysis was performed to quantify the implications of the difference between the vitals measured by the reference monitor and the PPG wristband. Clarke error grid analysis was originally developed for blood glucose measurements, and the boundaries of the different zones were adapted on the basis of the Modified Early Warning Score protocol used in our hospital [8,17,24,25].

Results

In total, 68 postsurgical patients were enrolled, of whom 6 were excluded from HR analysis due to either unavailable ECG reference (n=2) or a recording length of less than 15 minutes (n=4). For RR analysis, 14 patients were excluded from further analysis due to either lack of sufficient quality capnography reference data (n=9) or insufficient recording length (n=5) (Figure 1). The characteristics of the included population are shown in (Table 1).

Figure 1. Flowchart of patient inclusion for heart rate (HR) and respiratory rate (RR) analysis. A total of 62 patients were included in the data analysis, of whom 8 were only included in the HR analysis and 54 were included in both analyses. ECG: electrocardiography.



XSL•FO
RenderX

Table 1. Population demographics (N=62).

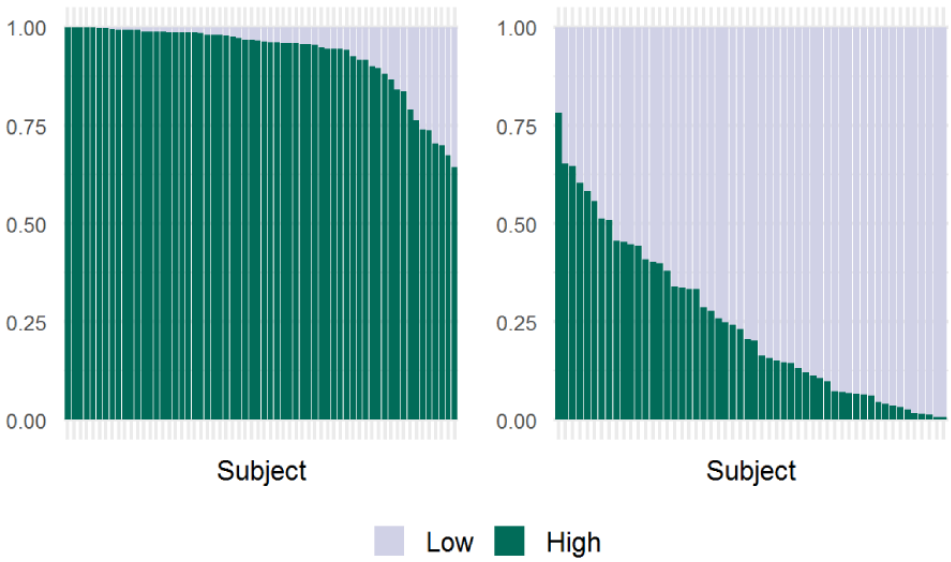
Variable	Value
Female, n (%)	33 (53)
Age (Years), mean (SD)	55 (15)
BMI (kg/m^2), median (IQR)	34 (25-40)
Surgery type, n (%)	
Gastric bypass	21 (34)
Gastric sleeve	9 (15)
Esophagectomy	7 (11)
Hyperthermic intraperitoneal chemotherapy	9 (15)
Pancreatectomy	4 (6)
Low anterior resection or abdominoperineal resection with intraoperative radiation therapy	6 (10)
Low anterior resection or abdominoperineal resection without intraoperative radiation therapy	4 (6)
Debulking	2 (3)
Duration of surgery (minutes), median (IQR)	144 (76-342)
Postoperative admission to the intensive care unit, n (%)	27 (43)

HR Assessment

For HR assessment, a total of 146 hours of data, from both the PACU or ICU patient monitor and the wearable sensor, were collected in 62 patients. Per patient, a median of 1.2 hours of data (range 16 minutes to 10 hours) were collected. Overall, 492,987 (94%) of the PPG wristband data points were of sufficient quality to be included in the analysis. As shown in (Figure 2), the percentage of sufficient-quality HR data per patient varied among patients, and a median of 96% (IQR 92%-99%) of high-quality HR data were obtained. The gaps without high-quality HR data ranged from a length of 1 second to 7.2 minutes, and 96% of the gaps were of <60 seconds.

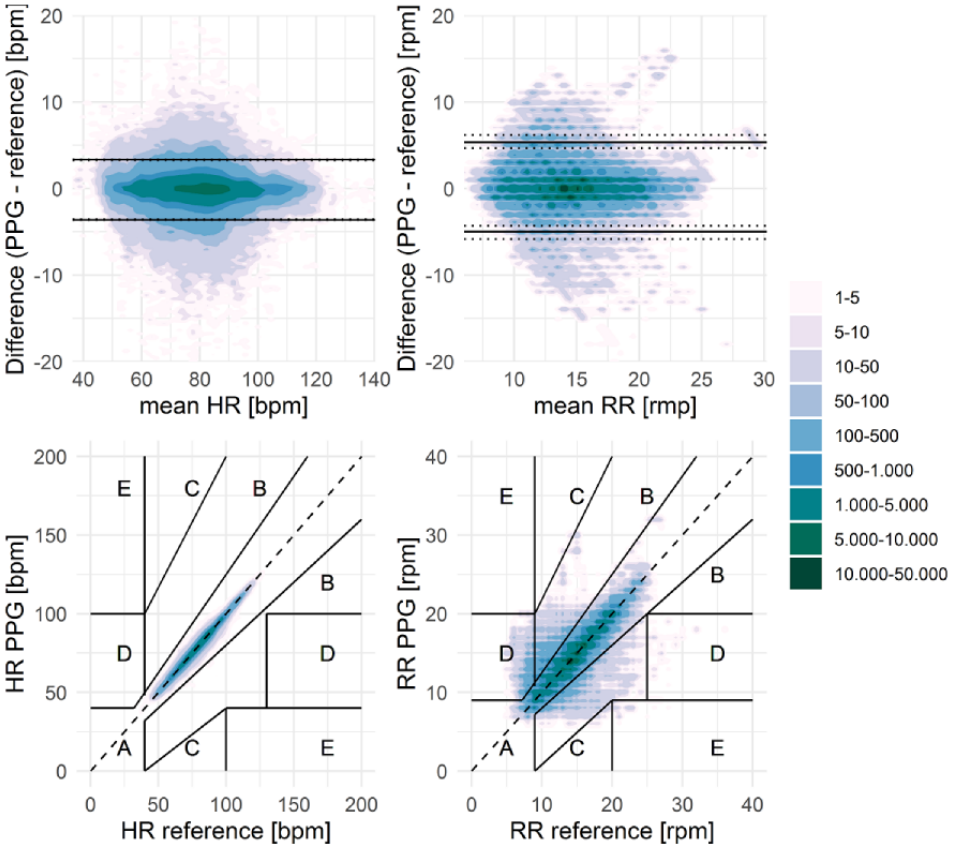
Bland-Altman and Clarke error grid analysis were used to assess the accuracy of the PPG wristband-measured HR (Figure 3 and Table 2). Bland-Altman analysis showed a bias of –0.15 bpm and limits of agreement of –3.62 to 3.32 bpm. As the limits of agreement lie within the predefined ≤5 bpm, the PPG wristband HR measurements met the required accuracy. Clarke error grid analysis showed that 100% (484,085 data points) of the measurements were within the clinically acceptable zones A and B, indicating that no incorrect treatment would result from the use of PPG wristband–derived HR values. Splitting the data on the basis of the unit patients were admitted to (ICU vs PACU) showed comparable availability of good-quality PPG measurements and similar bias and limits of agreement (Multimedia Appendix 1).

Figure 2. Availability of photoplethysmography wristband data of high-quality data for heart rate (left) and respiratory rate (right) expressed as the percentage of seconds with high- and low-quality data.



XSL•FO
RenderX

Figure 3. Bland-Altman (top) and Clarke error grid (bottom) plots of the vital parameters obtained from the photoplethysmography (PPG) wristband and reference monitor, each data point represents 1 second. The upper figures depict Bland-Altman analysis for heart rate (HR; left) and respiratory rate (RR; right). Limits of agreement are indicated by the black lines, dashed lines represent the 95% CIs of the limits of agreement. The bottom figures depict the Clarke error grid analysis for HR (left) and RR (right) comparing the measurements of the reference monitor (x-axis) and the PPG wristband (y-axis). Zone A represents data points that differ less than 20% from the reference or are correctly identified as bradycardia or bradypnea. Zone B represents data points that differ by more than 20% but would not cause unnecessary treatment. Zone C represents points that would lead to unnecessary treatment for patients with normal vital parameters. Zone D represents failure to detect bradycardia or bradypnea, or tachycardia or tachypnea. Zone E represents data points where bradycardia or bradypnea and tachycardia or tachypnea are confused.



XSL•FO
RenderX

Table 2. Agreement and clinical accuracy of heart rate (HR) and respiratory rate (RR) measured by the photoplethysmography (PPG) wristband compared to those of the reference monitor.

	HR	RR
Data availability		
Patients, n	62	54
Measurements, n	526,833	495,217
Good-quality PPG wristband, n (%)	492,987 (94)	170,383 (34)
Good-quality reference, n (%)	515,991 (98)	367,092 (74)
Both good quality, n (%)	484,096 (92)	128,816 (26)
Bland-Altman analysis		
Pearson *r*	0.99	0.80
Bias (bpm/rpm), mean (SD)	–0.15 (1.8)	0.17 (2.6)
Lower limit of agreement (bpm/rpm), lower limit (95% CI)	–3.62 (–3.7 to –3.6)	–4.99 (–5.8 to –4.3)
Upper limit of agreement (bpm/rpm), upper limit (95% CI)	3.32 (3.3 to 3.4)	5.33 (4.7 to 6.2)
Within 5 bpm or 3 rpm, %	98	93
Clarke error grid analysis, n (%)		
A	483,716 (99.9)	115,434 (89.6)
B	369 (0.1)	10,781 (8.4)
C	0 (0)	61 (0)
D	11 (0)	2499 (1.9)
E	0 (0)	41 (0)
A+B	484,085 (100)	126,215 (98.0)

RR Assessment

For RR, a total of 138 hours of data, from both the PACU or ICU patient monitor and the wearable sensor, were collected from among 54 patients. A median of 1.2 hours (range 16 minutes to 11 hours) of data were collected per patient. Overall, 170,383 (34%) of the PPG wristband RR measurements were of sufficient quality to be included in further analysis. Figure 2 shows the availability of high-quality data per patient, a median of 20% (IQR 7%-40%) of sufficient-quality RR data were obtained. The gaps without high-quality RR data ranged from a length of 1 second to 67 minutes, and 81% of the gaps were of <60 seconds.

Bland-Altman analysis of the PPG wristband–measured RR showed a bias of 0.17 rpm and limits of agreement of –4.99 to 5.33 rpm. As 93% of the RR measurements met the predefined ≤3 rpm, the limits of agreement were wider than the predefined ±3 rpm. Clarke error grid analysis showed that 98% of the data points were within the clinically acceptable zones A and B, indicating that the differences between the PPG wristband RR and reference monitor only have limited clinical implications. Most of the remaining 2% of data points lie within zone D, which indicates failure to detect impaired RR either due to failure to detect bradypnea (1.85%) or tachypnea (0.09%). Splitting the data based on the unit patients were admitted to (ICU vs PACU) showed a numerically higher percentage of available good-quality PPG measurements in the ICU and numerically wider limits of agreement in the PACU (Multimedia Appendix 1).

Discussion

Principal Findings

The use of wearable sensors to monitor hospitalized patients is rapidly attracting attention in the clinical community. However, prior to the introduction of these devices in clinical practice, their performance in the patient population of interest needs to be established. As postoperative patients are currently only monitored using spot checks for the duration for which they are in the general ward, this population could benefit from wearable monitoring. This study focused on the performance of a wearable PPG wristband for the measurement of HR and RR in postoperative patients.

For HR, the device was able to accurately measure the vital parameter as the bias and limits of agreement were within the predefined ≤5 bpm. Any differences between the PPG wristband and reference monitor were found to be clinically acceptable since 100% of the measurements were within zones A and B of the Clarke error grid. Additionally, the wearable PPG sensor would be feasible in terms of data availability for HR as the device only reported low quality for 4% of the HR measurements.

For RR, 93% of the included measurements were within the predefined ≤3 rpm and while the bias was within this threshold, the limits of agreement were wider than the predefined cutoff. However, as 98% of the included measurements lie within zones A and B of the Clarke error grid, the device does provide

XSL•FO
RenderX

clinically acceptable measurements for the large majority of the included measurements. The detection of RR by the wrist-worn PPG is easily corrupted by motion artifacts, leading to the exclusion of 66% of the measurements due to a low quality index. Therefore, the wearable PPG wristband would be unable to continuously measure RR. However, with 81% of the gaps of <1 minute, the device is able to measure RR more frequently than the current intermittent monitoring and therefore could potentially replace the RR measurements during the spot checks that are currently performed manually 3 times a day.

Comparisons to Prior Work

The accuracy of HR measurements by the same PPG wristband was previously studied in the PACU of our hospital. In the cohort of this study, the clinical accuracy in the PACU and ICU was confirmed with comparable results [15]. The accuracy of another wrist-worn PPG personal fitness tracker sensor for the monitoring of HR in hospitalized patients was previously studied by Kroll et al [12], who reported a bias of –4.7 and lower and upper limits of agreement of –31 and 21, respectively. Additionally, 73% of their measurements met the desired ≤5 bpm. Our findings with the ELAN PPG wristband show better agreement with the reference signal than their findings using the Fitbit Charge HR. The accuracy of another wrist-worn PPG sensor, the CardiacSense, in ambulatory patients was studied by Hochstadt et al [13]. As they reported their findings regarding the length of peak intervals rather than HR, comparison of results is difficult.

Limited data on the accuracy of RR measurement using PPG in clinical settings are available. Touw et al [26] studied the accuracy of finger-cuff PPG RR measurements in patients receiving procedural sedation and analgesia and found a bias of –2.0 rpm with limits of agreement from –12.4 to 8.4 rpm. Compared to their findings, the PPG wristband used in this study can measure RR with a smaller bias and smaller limits of agreement. Haveman et al [27] compared upper arm–worn wearable PPG measurements of HR and RR to manual those performed by nursing staff. They found a moderate relationship for HR and a poor relationship for RR. However, their results cannot be easily compared to ours as gold-standard measurements were unavailable in their cohort. Additionally, Haveman et al [28] described lower accuracy and data availability for upper arm–measured PPG RR during activity in volunteers. Patient activity level could therefore be a potential factor that relates to the differences between the RR in the ICU and the PACU. However, as the postoperative unit a patient is admitted to is chosen on the basis of surgery type, severity, and patient characteristics, this trial does not allow drawing conclusions regarding the origin of these differences. Papini et al [29] studied respiratory activity in a sleep-disordered population using a wrist-worn PPG device. They found a median correlation of 0.62 and a median per patient coverage of 75.3%. Comparison of the accuracy to our findings is complicated as we reported a correlation over the entire data set; however, an overall correlation of 0.80 in this study indicates a better agreement between the 2 RR measurements. However, their median per-patient coverage of 75.3% clearly outperforms the 20% found in the present population.

Strengths and Limitations

This analysis was performed in a real-world, clinically relevant patient population, as postoperative patients could benefit from wearable monitoring in low-acuity care settings such as the surgical ward. However, this study had some limitations. First, while capnography is the gold standard for RR monitoring, a good-quality reference RR signal could not be obtained for 9 patients, and for the patients who could be included, 26% of the capnography data had insufficient quality to be included. Second, the data for this trial were collected in the PACU and ICU rather than the general ward. However, we believe that our findings could reasonably be transferred to the general ward as patients became alert and mobile during their stay in these recovery units. Third, the analysis of trending ability of the device could be an interesting addition to the data analysis and can be included in future research if longer monitoring times of both the wearable and reference monitor are available.

Future Directions

Other potential future clinical applications of PPG wearables include the measurement of activity level, blood pressure, HR variability, energy expenditure, and the detection of atrial fibrillation [21,30-33]. In future clinical use, PPG wristbands thus have the potential to provide information on even more aspects of the patients' health status. This study shows that the ELAN PPG wristband can continuously measure HR with clinically acceptable accuracy. For RR, the device can perform clinically accurate measurements, but, due to limited coverage, can only be used to perform intermittent measurements.

Conclusions

The wearable PPG wristband can measure HR accurately and with sufficient coverage in postoperative patients. For RR, the large majority of the included data were clinically acceptable; however, the coverage of sufficient-quality RR measurements was low. Therefore, the PPG wristband would be able to perform continuous monitoring of HR and also report RR when sufficient-quality measurements are obtained. Before implementing such PPG-based wearable devices in clinical practice, both accuracy and coverage should be considered.

Acknowledgments

The authors would like to thank Linda Eerikainen, Koen de Groot, and Wim Verhaegh for their contribution to the data preparation. The trial was partly financed by a grant from the Rijksdienst voor Ondernemend Nederland (grant RVO ITEA 161006).

XSL•FO
RenderX

Conflicts of Interest

RAB acts as a clinical consultant for Philips Research. AGB is an employee of Philips Research. HMdM was an employee of Philips Research at the time this study was conducted.

Multimedia Appendix 1
Results split based on unit the patients were admitted to postoperatively.
[PDF File (Adobe PDF File), 729 KB - periop_v6i1e40474_app1.pdf]

References

1. Buist M, Bernard S, Nguyen TV, Moore G, Anderson J. Association between clinically abnormal observations and subsequent in-hospital mortality: a prospective study. Resuscitation 2004 Aug;62(2):137-141. [doi: 10.1016/j.resuscitation.2004.03.005] [Medline: 15294398]
2. Hillman KM, Bristow PJ, Chey T, Daffurn K, Jacques T, Norman SL, et al. Antecedents to hospital deaths. Intern Med J 2001 Aug;31(6):343-348. [doi: 10.1046/j.1445-5994.2001.00077.x] [Medline: 11529588]
3. Hillman KM, Bristow PJ, Chey T, Daffurn K, Jacques T, Norman SL, et al. Duration of life-threatening antecedents prior to intensive care admission. Intensive Care Med 2002 Nov;28(11):1629-1634. [doi: 10.1007/s00134-002-1496-y] [Medline: 12415452]
4. Nurmi J, Harjola VP, Nolan J, Castrén M. Observations and warning signs prior to cardiac arrest. Should a medical emergency team intervene earlier? Acta Anaesthesiol Scand 2005 May;49(5):702-706. [doi: 10.1111/j.1399-6576.2005.00679.x] [Medline: 15836688]
5. Fieselmann JF, Hendryx MS, Helms CM, Wakefield DS. Respiratory rate predicts cardiopulmonary arrest for internal medicine inpatients. J Gen Intern Med 1993 Jul;8(7):354-360. [doi: 10.1007/BF02600071] [Medline: 8410395]
6. Walston JM, Cabrera D, Bellew SD, Olive MN, Lohse CM, Bellolio MF. Vital signs predict rapid-response team activation within twelve hours of emergency department admission. West J Emerg Med 2016 May;17(3):324-330 [FREE Full text] [doi: 10.5811/westjem.2016.2.28501] [Medline: 27330665]
7. Cuthbertson BH, Boroujerdi M, McKie L, Aucott L, Prescott G. Can physiological variables and early warning scoring systems allow early recognition of the deteriorating surgical patient? Crit Care Med 2007 Feb;35(2):402-409. [doi: 10.1097/01.CCM.0000254826.10520.87] [Medline: 17205002]
8. Subbe CP, Kruger M, Rutherford P, Gemmel L. Validation of a modified Early Warning Score in medical admissions. QJM 2001 Oct;94(10):521-526. [doi: 10.1093/qjmed/94.10.521] [Medline: 11588210]
9. Chen J, Hillman K, Bellomo R, Flabouris A, Finfer S, Cretikos M, MERIT Study Investigators for the Simpson Centre, ANZICS Clinical Trials Group. The impact of introducing medical emergency team system on the documentations of vital signs. Resuscitation 2009 Jan;80(1):35-43. [doi: 10.1016/j.resuscitation.2008.10.009] [Medline: 19010579]
10. Cretikos MA, Bellomo R, Hillman K, Chen J, Finfer S, Flabouris A. Respiratory rate: the neglected vital sign. Med J Aust 2008 Jun 02;188(11):657-659. [doi: 10.5694/j.1326-5377.2008.tb01825.x] [Medline: 18513176]
11. Suaste-Gómez E, Hernández-Rivera D, Sánchez-Sánchez AS, Villarreal-Calva E. Electrically insulated sensing of respiratory rate and heartbeat using optical fibers. Sensors (Basel) 2014 Nov 14;14(11):21523-21534 [FREE Full text] [doi: 10.3390/s141121523] [Medline: 25405510]
12. Kroll RR, Boyd JG, Maslove DM. Accuracy of a wrist-worn wearable device for monitoring heart rates in hospital inpatients: a prospective observational study. J Med Internet Res 2016 Sep 20;18(9):e253 [FREE Full text] [doi: 10.2196/jmir.6025] [Medline: 27651304]
13. Hochstadt A, Havakuk O, Chorin E, Schwartz AL, Merdler I, Laufer M, et al. Continuous heart rhythm monitoring using mobile photoplethysmography in ambulatory patients. J Electrocardiol 2020;60:138-141. [doi: 10.1016/j.jelectrocard.2020.04.017] [Medline: 32361522]
14. Koshy AN, Sajeev JK, Nerlekar N, Brown AJ, Rajakariar K, Zureik M, et al. Utility of photoplethysmography for heart rate estimation among inpatients. Intern Med J 2018 May;48(5):587-591. [doi: 10.1111/imj.13777] [Medline: 29722189]
15. Mestrom E, Deneer R, Bonomi AG, Margarito J, Gelissen J, Haakma R, et al. Validation of heart rate extracted from wrist-based photoplethysmography in the perioperative setting: prospective observational study. JMIR Cardio 2021 Nov 04;5(2):e27765 [FREE Full text] [doi: 10.2196/27765] [Medline: 34734834]
16. Churpek MM, Yuen TC, Winslow C, Meltzer DO, Kattan MW, Edelson DP. Multicenter comparison of machine learning methods and conventional regression for predicting clinical deterioration on the wards. Crit Care Med 2016 Feb;44(2):368-374 [FREE Full text] [doi: 10.1097/CCM.0000000000001571] [Medline: 26771782]
17. van der Stam JA, Mestrom EHJ, Nienhuijs SW, de Hingh IHJT, Boer A, van Riel NAW, et al. A wearable patch based remote early warning score (REWS) in major abdominal cancer surgery patients. Eur J Surg Oncol 2023 Jan;49(1):278-284 [FREE Full text] [doi: 10.1016/j.ejso.2022.08.034] [Medline: 36085116]
18. van Ede ES, Scheerhoorn J, Bonomi AG, Buise MP, Bouwman RA, Nienhuijs SW. Continuous remote monitoring in post-bariatric surgery patients: development of an early warning protocol. Surg Obes Relat Dis 2022 Nov;18(11):1298-1303 [FREE Full text] [doi: 10.1016/j.soard.2022.06.018] [Medline: 35850957]

XSL•FO
RenderX

19. Papini GB, Fonseca P, van Gilst MM, Bergmans JWM, Vullings R, Overeem S. Wearable monitoring of sleep-disordered breathing: estimation of the apnea-hypopnea index using wrist-worn reflective photoplethysmography. Sci Rep 2020 Aug 11;10(1):13512 [FREE Full text] [doi: 10.1038/s41598-020-69935-7] [Medline: 32782313]
20. Kathirvel P, Sabarimalai Manikandan M, Prasanna SRM, Soman KP. An efficient R-peak detection based on new nonlinear transformation and first-order Gaussian differentiator. Cardiovasc Eng Tech 2011 Oct 12;2(4):408-425. [doi: 10.1007/s13239-011-0065-3]
21. Bonomi AG, Schipper F, Eerikäinen LM, Margarito J, van Dinther R, Muesch G, et al. Atrial fibrillation detection using a novel cardiac ambulatory monitor based on photo-plethysmography at the wrist. J Am Heart Assoc 2018 Aug 07;7(15):e009351 [FREE Full text] [doi: 10.1161/JAHA.118.009351] [Medline: 30371247]
22. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986 Feb 08;1(8476):307-310. [Medline: 2868172]
23. Zou GY. Confidence interval estimation for the Bland-Altman limits of agreement with multiple observations per individual. Stat Methods Med Res 2013 Dec;22(6):630-642. [doi: 10.1177/0962280211402548] [Medline: 21705434]
24. Clarke WL, Cox D, Gonder-Frederick LA, Carter W, Pohl SL. Evaluating clinical accuracy of systems for self-monitoring of blood glucose. Diabetes Care 1987;10(5):622-628. [doi: 10.2337/diacare.10.5.622] [Medline: 3677983]
25. Breteler MJM, KleinJan EJ, Dohmen DAJ, Leenen LPH, van Hillegersberg R, Ruurda JP, et al. Vital signs monitoring with wearable sensors in high-risk surgical patients: a clinical validation study. Anesthesiology 2020 Mar;132(3):424-439 [FREE Full text] [doi: 10.1097/ALN.0000000000003029] [Medline: 31743149]
26. Touw HRW, Verheul MH, Tuinman PR, Smit J, Thöne D, Schober P, et al. Photoplethysmography respiratory rate monitoring in patients receiving procedural sedation and analgesia for upper gastrointestinal endoscopy. J Clin Monit Comput 2017 Aug;31(4):747-754 [FREE Full text] [doi: 10.1007/s10877-016-9890-0] [Medline: 27236561]
27. Haveman ME, van Melzen R, Schuurmann RCL, El Moumni M, Hermens HJ, Tabak M, et al. Continuous monitoring of vital signs with the Everion biosensor on the surgical ward: a clinical validation study. Expert Rev Med Devices 2021 Dec;18(sup1):145-152. [doi: 10.1080/17434440.2021.2019014] [Medline: 34937478]
28. Haveman ME, van Rossum MC, Vaseur RME, van der Riet C, Schuurmann RCL, Hermens HJ, et al. Continuous monitoring of vital signs with wearable sensors during daily life activities: validation study. JMIR Form Res 2022 Jan 07;6(1):e30863 [FREE Full text] [doi: 10.2196/30863] [Medline: 34994703]
29. Papini GB, Fonseca P, Gilst MMV, Bergmans JW, Vullings R, Overeem S. Respiratory activity extracted from wrist-worn reflective photoplethysmography in a sleep-disordered population. Physiol Meas 2020 Jul 08;41(6):065010. [doi: 10.1088/1361-6579/ab9481] [Medline: 32428875]
30. Rachim VP, Chung W. Multimodal wrist biosensor for wearable cuff-less blood pressure monitoring system. Sci Rep 2019 May 28;9(1):7947 [FREE Full text] [doi: 10.1038/s41598-019-44348-3] [Medline: 31138845]
31. Moraes J, Rocha M, Vasconcelos G, Vasconcelos Filho J, de Albuquerque V, Alexandria A. Advances in photopletysmography signal analysis for biomedical applications. Sensors (Basel) 2018 Jun 09;18(6):1894 [FREE Full text] [doi: 10.3390/s18061894] [Medline: 29890749]
32. Hochstadt A, Chorin E, Viskin S, Schwartz AL, Lubman N, Rosso R. Continuous heart rate monitoring for automatic detection of atrial fibrillation with novel bio-sensing technology. J Electrocardiol 2019;52:23-27. [doi: 10.1016/j.jelectrocard.2018.10.096] [Medline: 30476634]
33. Wallen MP, Gomersall SR, Keating SE, Wisløff U, Coombes JS. Accuracy of heart rate watches: implications for weight management. PLoS One 2016;11(5):e0154420 [FREE Full text] [doi: 10.1371/journal.pone.0154420] [Medline: 27232714]

Abbreviations

ECG: electrocardiography
HR: heart rate
ICU: intensive care unit
PACU: postanesthesia care unit
PPG: photoplethysmography
RR: respiratory rate

XSL•FO
RenderX

Edited by T Leung; submitted 23.06.22; peer-reviewed by M Haveman, V Gupta; comments to author 25.11.22; revised version received 12.01.23; accepted 31.01.23; published 20.02.23.

Please cite as:
van der Stam JA, Mestrom EHJ, Scheerhoorn J, Jacobs FENB, Nienhuijs S, Boer AK, van Riel NAW, de Morree HM, Bonomi AG, Scharnhorst V, Bouwman RA
The Accuracy of Wrist-Worn Photoplethysmogram–Measured Heart and Respiratory Rates in Abdominal Surgery Patients: Observational Prospective Clinical Validation Study
JMIR Perioper Med 2023;6:e40474
URL: https://periop.jmir.org/2023/1/e40474
doi:10.2196/40474
PMID:36804173

©Jonna A van der Stam, Eveline H J Mestrom, Jai Scheerhoorn, Fleur E N B Jacobs, Simon Nienhuijs, Arjen-Kars Boer, Natal A W van Riel, Helma M de Morree, Alberto G Bonomi, Volkher Scharnhorst, R Arthur Bouwman. Originally published in JMIR Perioperative Medicine (http://periop.jmir.org), 20.02.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Perioperative Medicine, is properly cited. The complete bibliographic information, a link to the original publication on http://periop.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Prediction of Pelvic Organ Prolapse Postsurgical Outcome Using Biomaterial-Induced Blood Cytokine Levels: Machine Learning Approach

Mihyun Lim Waugh[1], BS; Nicholas Boltin[1], PhD; Lauren Wolf[1], PhD; Jane Goodwin[2], BS; Patti Parker[3], BSN; Ronnie Horner[4], PhD; Matthew Hermes[5], PhD; Thomas Wheeler[6], MD; Richard Goodwin[2], PhD; Melissa Moss[1,7], PhD

[1]Biomedical Engineering Program, University of South Carolina, Columbia, SC, United States
[2]Biomedical Sciences, University of South Carolina School of Medicine - Greenville, Greenville, SC, United States
[3]Prisma Health Greenville, Greenville, SC, United States
[4]Department of Health Services Research and Administration, University of Nebraska Medical Center, Omaha, NE, United States
[5]Department of Bioengineering, Clemson University, Clemson, SC, United States
[6]Spartanburg Regional Healthcare System, Spartanburg, SC, United States
[7]Department of Chemical Engineering, University of South Carolina, Columbia, SC, United States

Corresponding Author:
Melissa Moss, PhD
Department of Chemical Engineering, University of South Carolina
301 Main St.
Rm 2C02
Columbia, SC, 29208
United States
Phone: 1 803 777 5604
Email: mossme@cec.sc.edu

Abstract

Background: Pelvic organ prolapse (POP) refers to symptomatic descent of the vaginal wall. To reduce surgical failure rates, surgical correction can be augmented with the insertion of polypropylene mesh. This benefit is offset by the risk of mesh complication, predominantly mesh exposure through the vaginal wall. If mesh placement is under consideration as part of prolapse repair, patient selection and counseling would benefit from the prediction of mesh exposure; yet, no such reliable preoperative method currently exists. Past studies indicate that inflammation and associated cytokine release is correlated with mesh complication. While some degree of mesh-induced cytokine response accompanies implantation, excessive or persistent cytokine responses may elicit inflammation and implant rejection.

Objective: Here, we explore the levels of biomaterial-induced blood cytokines from patients who have undergone POP repair surgery to (1) identify correlations among cytokine expression and (2) predict postsurgical mesh exposure through the vaginal wall.

Methods: Blood samples from 20 female patients who previously underwent surgical intervention with transvaginal placement of polypropylene mesh to correct POP were collected for the study. These included 10 who experienced postsurgical mesh exposure through the vaginal wall and 10 who did not. Blood samples incubated with inflammatory agent lipopolysaccharide, with sterile polypropylene mesh, or alone were analyzed for plasma levels of 13 proinflammatory and anti-inflammatory cytokines using multiplex assay. Data were analyzed by principal component analysis (PCA) to uncover associations among cytokines and identify cytokine patterns that correlate with postsurgical mesh exposure through the vaginal wall. Supervised machine learning models were created to predict the presence or absence of mesh exposure and probe the number of cytokine measurements required for effective predictions.

Results: PCA revealed that proinflammatory cytokines interferon gamma, interleukin 12p70, and interleukin 2 are the largest contributors to the variance explained in PC 1, while anti-inflammatory cytokines interleukins 10, 4, and 6 are the largest contributors to the variance explained in PC 2. Additionally, PCA distinguished cytokine correlations that implicate prospective therapies to improve postsurgical outcomes. Among machine learning models trained with all 13 cytokines, the artificial neural network, the highest performing model, predicted POP surgical outcomes with 83% (15/18) accuracy; the same model predicted

XSL•FO
RenderX

POP surgical outcomes with 78% (14/18) accuracy when trained with just 7 cytokines, demonstrating retention of predictive capability using a smaller cytokine group.

Conclusions: This preliminary study, incorporating a sample size of just 20 participants, identified correlations among cytokines and demonstrated the potential of this novel approach to predict mesh exposure through the vaginal wall following transvaginal POP repair surgery. Further study with a larger sample size will be pursued to confirm these results. If corroborated, this method could provide a personalized medicine approach to assist surgeons in their recommendation of POP repair surgeries with minimal potential for adverse outcomes.

(JMIR Perioper Med 2023;6:e40402) doi:10.2196/40402

KEYWORDS

pelvic organ prolapse; polypropylene mesh; inflammatory response; cytokines; principal component analysis; supervised machine learning models; surgical outcome prediction; biomaterial; repair surgery

Introduction

Pelvic organ prolapse (POP), defined as symptomatic descent of the vagina and surrounding pelvic organs, affects approximately 50% of parous women and 6% of nonparous women between ages 20 and 59 years [1], with almost 300,000 POP surgeries performed per year [2]. To reduce anatomical recurrence, surgical treatment may include the insertion of polypropylene mesh into the vaginal wall to provide mechanical support and reinforcement of the prolapsed organs. Unfortunately, postsurgical mesh complication, predominantly mesh exposure through the vaginal wall, occurs with some frequency and results in decreased quality of life, leaving patients with costly residual symptoms and emotional distress [3]. Patients may elect for surgical reintervention to revise or remove the mesh implantation. In fact, according to Reid et al [4], 37 (8%) out of 482 patients underwent further surgery to remove the mesh, and 7 (2%) patients repeated the prolapse surgery. These complications provoked the removal of transvaginal mesh kits from the market by the Food and Drug Administration in 2019. A clinical decision support tool to better inform both patients and surgeons about the risk of complications following POP surgery may allow for the reintroduction of this advantageous surgical augmentation.

Inflammatory responses are associated with mesh exposure due to asymptomatic mesh infection that inhibits the mesh from integrating with the surrounding environment [5]. While some degree of mesh-induced cytokine response is necessary for successful implantation, excess or unattenuated cytokine response could result in chronic inflammation and implant rejection. As chronic inflammation progresses, granulation tissues formed during the foreign body reaction will evolve into mesh encapsulation by regular dense connective tissue and myofibroblast-induced contracture around the implant, which can result in mesh exposure [6,7]. The balance between proinflammatory and anti-inflammatory agents is critical in achieving successful mesh implantation, and this balance may be influenced by the individual's response to the implant material. Thus, leveraging a patient's immune response to the biomaterial could facilitate the prediction of postsurgical outcomes.

Leveraging a patient-specific, multifaceted immune response for the prediction of postsurgical complications is an ideal problem for the application of principal component analysis (PCA) and supervised machine learning models. In fact, this approach has been used to predict complications following other surgical procedures as well as progressive disease outcomes. In a liver transplant study, Raji and Vinod Chandra [8] applied PCA to a composite medical data set comprised of donors' medical information as well as the recipients' medical history and implemented an artificial neural network to predict the long-term survival of liver transplant patients. In an oral cancer retrospective study by Chu et al [9], PCA along with bivariate analyses were used to highlight correlated variables from the patient data, which included patient demographics and clinicopathological tumor data (including tumor sites, disease staging, etc), and to predict oral cancer progress.

PCA and supervised machine learning have also been applied to biological measurements for predicting medical outcomes. Tseng et al [10] built a predictive model for cardiac surgery–associated acute kidney injury (AKI) using preoperative biochemistry data in combination with patient demographic characteristics and clinical condition. Incorporating a different type of biological measurement, a glioblastoma study by Akbari et al [11] used PCA and support vector machines to distinguish multiparametric magnetic resonance imaging signatures and quantify the patterns to predict regions of tumor recurrence after surgery. Chen et al [12] demonstrated that specifically including immune data in predictive models enhances predictive capacity. These researchers implemented machine learning models using individual patient immune data, such as blood cytokine levels, to predict severe AKI after cardiac surgery and found that this approach provided a far superior prediction tool compared to a clinical factor–based model [12].

The application of PCA and machine learning to predict postsurgical complications in women after POP surgery has also shown promising results. In the study of Jelovsek et al [13], statistical modeling uses 32 candidate risk factors (ie, age, race, smoking history, etc) identified by consensus with surgical outcomes to predict postsurgical complications. This approach of using personalized preoperative decision-making based on the individual's medical history presents a better predictive model to postsurgical complications and offers a more effective decision support tool than the practice of counseling patients using average success rates reported from large, randomized studies [13]. However, this predictive method does not leverage the patient's potential immune response to the surgery involving polypropylene mesh.

XSL•FO
RenderX

In this preliminary study, we explored the levels of baseline and stimulus-induced cytokines in blood isolated from patients who had undergone POP repair surgery with a polypropylene mesh. Proinflammatory and anti-inflammatory cytokine levels from these data were analyzed using PCA to establish the principal components (PCs) and to identify associative or opposing trends among cytokines. In addition, supervised machine learning models were applied to demonstrate predictive capabilities when models were trained with either all 13 cytokines or a smaller group of 7 cytokines determined most effective by a random forest method. The results demonstrate that leveraging PCA and supervised machine learning models to predict outcomes of vaginal mesh implantation has the potential to benefit future patients when they are faced with this surgical decision, which carries a relatively high risk of unsuccessful surgical outcome.

Methods

Study Population

In total, 20 healthy, nonpregnant female participants aged 56-89 years at Prisma Health Greenville Memorial Hospital with a history of surgical intervention to correct POP via a procedure that used polypropylene mesh were selected for the study. The participants, who were not matched, included 10 who experienced postsurgical mesh complication in which the implanted mesh protruded through the vaginal wall (also referred to as mesh exposure) and 10 participants who did not experience this complication post surgery. This sample size was estimated as an effective cohort for the pilot study using a 1-tailed *t* test based on an a priori power analysis, which indicates the number of patients for a given theoretical minimum study power as a function of the expected difference between patients with and those without mesh exposure, or the Cohen *d*. We assumed a conservative study power of 0.80 and a 100%, or 2-fold, difference in the level of a given cytokine between individuals with and without mesh exposure, equivalent to a Cohen *d* of 1. Here, 20 patients with equal distribution among the 2 groups are needed to observe the difference with a probability of .1. Participants with POP recurrence or taking medications that would alter inflammatory response were excluded from this study.

Ethics Approval

The study protocol was approved by the institutional review board (IRB) of Prisma Health (Pro00067964). Informed consent from all study participants was obtained using an IRB-approved informed consent form. All samples collected and data analyzed were deidentified and followed IRB protocol.

Blood Sample Collection and Processing

Blood samples were obtained from the 20 selected participants. Approximately 12 mL of blood was drawn from the upper extremity of each participant into 3 BD Vacutainer EDTA-coated tubes. Deidentified blood samples were then transferred on ice to a laboratory facility at the University of South Carolina School of Medicine Greenville for immediate processing. Each participant's blood sample was divided into equal aliquots for 24-hour incubation at 37 °C under 3 distinct conditions: (1) incubation with inflammatory agent lipopolysaccharide (LPS) at 20 ng/mL (positive control), (2) incubation with sterile polypropylene mesh area of 2 cm × 2 cm (experimental), and (3) incubation alone (negative control). After incubation, the plasma layer was collected following centrifugation (1500 × g, 10 min, 4 °C) and immediately stored at –80 °C.

Measurement of Blood Cytokine Levels

Cytokine levels in each blood sample were quantified using the bead-based MILLIPLEX Human Cytokine/Chemokine/Growth Factor Panel A—Immunology Multiplex Assay (EMD Millipore Corp), which is composed of analytes for target cytokines interleukin 1α (IL-1α), IL-1β, IL-2, IL-4, IL-6, IL-8, IL-10, IL-12p40, IL-12p70, IL-17A, interferon-gamma (IFN-γ, tumor necrosis factor-alpha (TNF-α), and granulocyte-macrophage colony-stimulating factor (GM-CSF). Frozen plasma samples were thawed at room temperature and analyzed following Milliplex protocol guidelines. Cytokine concentrations were measured using a Bio-Plex 200 (Bio-Rad) and Bio-Plex Manager software (Bio-Rad). Sample volume was doubled to ensure measurable levels of cytokines, and assay output data were adjusted to reflect concentrations in plasma samples. Each multiplex assay was performed in duplicate, and cytokine levels were evaluated in 3 independent measurements.

Data Analysis

Overview

Cytokine data gathered from the multiplex immunoassay were analyzed using data mining and predictive analytical methods. PCA was used to identify important cytokines by studying their contributions to each PC as well as to discern associations between cytokines. Supervised machine learning models were created to determine whether cytokine levels can accurately predict which patients are more likely to experience mesh exposure post surgery.

Descriptive Analytics

The statistical programming language R (version 4.1.2; R Foundation) was used to analyze raw cytokine data values generated from the multiplex immunoassay. The imported data structure contained 60 observations (20 participants × 3 independent measurements) and 40 total variable fields (13 cytokines × 3 blood treatments + 1 target variable). The target variable was the participant's outcome, which indicated a postsurgical complication that participants might have experienced following POP surgery. Observations marked "presence" represent participants who experienced mesh exposure through the vaginal wall. Observations marked "absence" represent participants who did not experience any mesh exposure through the vaginal wall. Univariate and multivariate methods were used to explore the data set, including identifying missing values, analyzing outliers, and visualizing frequency distributions.

PCA

PCA was performed using the FactoMineR package (version 2.4; R Foundation) [14] to identify associations between cytokines [15]. Before analysis, data transformations were

XSL•FO
RenderX

performed on each variable to correct for skewness in the distribution. The amount of skewness was calculated to assess the symmetry of distribution for each variable using equation 1, where is the sample mean and x_i and n are the individual observations and number of observations, respectively, within the sample [16]. Each cytokine's distribution was corrected for skewness using either a natural logarithm, square root, or inverse square root method. Using equation 2, z-score standardization was also applied to scale each cytokine variable, thus ensuring that the mean was equal to 0 and SD equal to 1. Biplots were created to visualize PCs with the highest degree of variance explained. The eigenvectors were overlayed on the biplots to visualize correlations and identify hidden patterns between cytokines.

Predictive Analytics

Supervised machine learning models were created using the caret package (version 6.0-90; R Foundation) [17] in the R programming language. The 4 models trained were decision tree, logistic regression, Naive Bayes, and artificial neural network. This approach focused on the data set from the experimental group only (cytokine expression for blood incubated with polypropylene mesh). Prior to creating the predictive models, the original data were split using an industry standard of 70% for training and 30% for testing. Each group contained an equal distribution of participants who did or did not experience postsurgical mesh exposure through the vaginal wall—the prediction target for each model. Each model was then trained using the 70% (42/60) subset and a cross-validation training control. A 10-fold cross-validation with 25% (15/60) left out replicated 3 times was used on each model to avoid bias and overfitting. From this, training accuracies are reported. Additional testing was performed for the prediction accuracy of each model using the 30% (18/60) test data. Prediction accuracies are reported along with sensitivity and specificity for the prediction of participants to experience postsurgical mesh exposure.

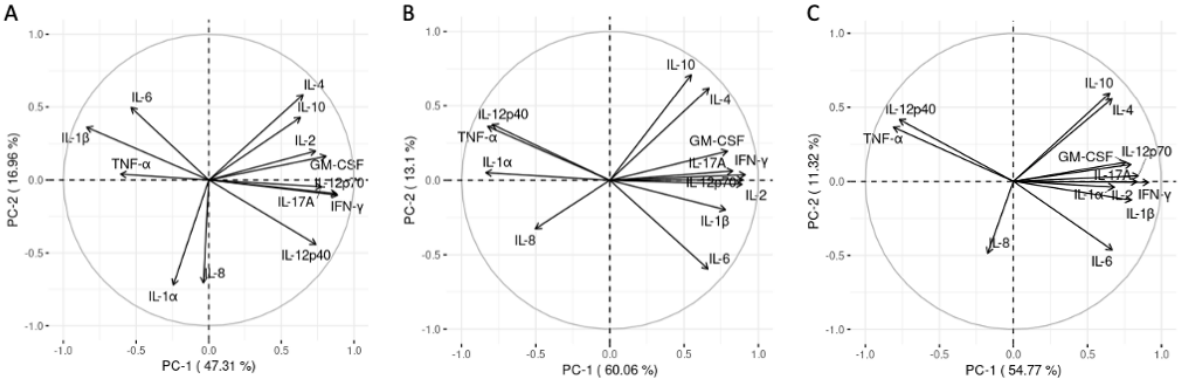
Additionally, this process was replicated to study the effects of reducing the number of cytokines needed to predict a postsurgical mesh exposure. A random forest algorithm was used to select important cytokines for this study. These models were trained and tested for accuracy, sensitivity, and specificity as detailed above. The results are compared to models trained with all 13 cytokines.

Results

PCA

To identify significant associations among the cytokines, PCA was used to examine a total of 60 blood samples (20 participants × 3 blood treatments). Among the 20 participants, 10 experienced postsurgical mesh exposure through the vaginal wall and 10 did not. Figure 1 depicts a biplot of each blood treatment and summarizes the intercorrelated relationships among individual inflammatory mediators. The combined variances explained for PC 1 and PC 2 in blood samples incubated with LPS (Figure 1A), polypropylene mesh (Figure 1B), or alone (Figure 1C) were approximately 64%, 73%, and 66%, respectively. In all 3 treatment groups, IL-10 and IL-4 align in the same directions, as do IL-12p70 and IFN-γ, indicating a positive correlation for both of these cytokine pairs. In contrast, IL-6 and IL-12p40 were negatively correlated when comparing stimulation of blood via LPS (Figure 1A) versus polypropylene mesh (Figure 1B). Only IL-1α displayed a negative correlation when comparing blood incubated with polypropylene mesh (Figure 1B) versus blood incubated alone (Figure 1C).

Figure 1. PCA was performed using cytokine levels in blood samples of postsurgical POP subjects; the analysis included 60 blood samples (20 subjects × 3 blood treatments), wherein each sample was evaluated in 3 independent measurements performed in duplicate. Biplots illustrating individual cytokines were constructed for blood samples incubated in the presence of LPS (A), incubated in the presence of polypropylene mesh (B), or incubated alone (C). Arrow direction indicates the cytokine correlation; arrow length indicates the magnitude of the variation. IL-1α: interleukin-1 alpha; IL-1β: interleukin-1 beta; IL-2: interleukin-2; IL-4: interleukin-4; IL-6: interleukin-6; IL-8: interleukin-8; IL-10: interleukin-10; IL-12p40: interleukin-12p40; IL-12 p70: interleukin-12p70; IL-17A: interleukin-17A; IFN-γ: interferon gamma; TNF-α: tumor necrosis factor-alpha; GM-CSF: granulocyte-macrophage colony-stimulating factor; PC: principal component.



When PCA was used to examine only blood samples incubated with polypropylene mesh, PC 1 and PC 2 explained 60.1% and 13.1% of the total data variance, respectively (Figure 1B). Figure 2A displays each cytokine's contribution to PC 1 and illustrates that IFN-γ, IL-12p70, and IL-2 are the predominant contributors to the variance explained in PC 1. In addition, IL-1α, IL-17A,

XSL•FO
RenderX

and TNF-α exhibited contributions above a level expected if the contributions were uniform. All other cytokines have contributions to PC 1 similar to or less than what would be expected if the contributions of all cytokines were uniform. Figure 2B illustrates that the predominant contributors to the variance explained in PC 2 are IL-10, IL-4, and IL-6. All other cytokines have contributions to PC 2 similar to or less than what would be expected if the contribution of all cytokines were uniform.

In order to visualize associations between the participants presenting the absence or presence of postsurgical mesh exposure through the vaginal wall, a biplot illustrating individual participants was created (Figure 3). This biplot reveals a high percentage of variability represented by the first 2 PCs (79.1%). Blood samples from participants who did not experience postsurgical mesh exposure were heavily represented by positive PC 1 values, while blood samples from participants with the presence of postsurgical mesh exposure were generally represented by positive PC 2 values.

Figure 2. PCA was performed using cytokine levels in patient blood samples incubated with polypropylene mesh; the analysis included 20 blood samples (20 subjects × 1 blood treatment), wherein each sample was evaluated in 3 independent measurements performed in duplicate. Each cytokine's contribution to PC 1 (A) and PC 2 (B) was determined. The dashed line at 7% corresponds to the expected value if the contribution were uniform. IL-1α: interleukin-1 alpha; IL-1β: interleukin-1 beta; IL-2: interleukin-2; IL-4: interleukin-4; IL-6: interleukin-6; IL-8: interleukin-8; IL-10: interleukin-10; IL-12p40: interleukin-12p40; IL-12p70: interleukin-12p70; IL-17A: interleukin-17A; IFN-γ: interferon gamma; TNF-α: tumor necrosis factor-alpha; GM-CSF: granulocyte-macrophage colony-stimulating factor; PC: principal component.

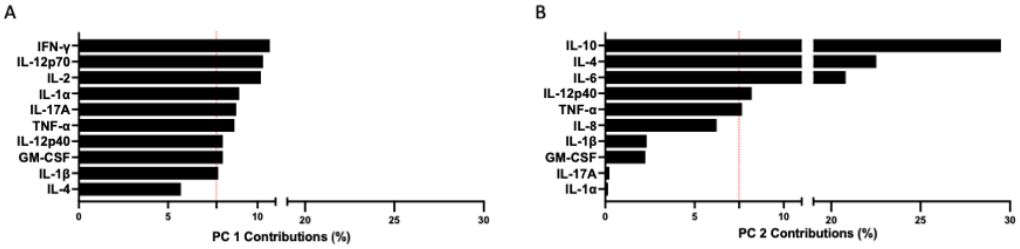
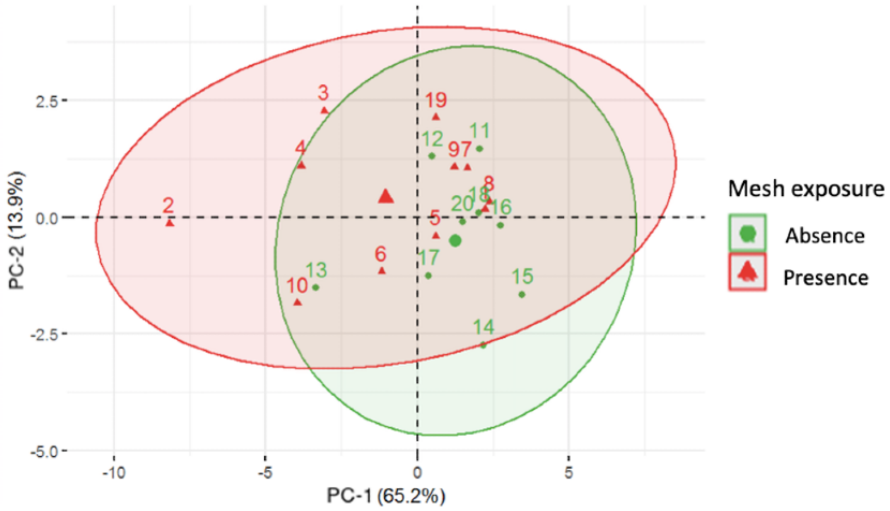


Figure 3. PCA was performed using cytokine levels in patient blood samples incubated with polypropylene mesh; the analysis included 20 blood samples (20 subjects × 1 blood treatment), wherein each sample is represented by the average of 3 independent measurements performed in duplicate. A biplot was constructed illustrating individual patient averages (indicated by numbers) exhibiting the presence (red triangle) or absence (green circle) of mesh exposure through the vaginal wall. Concentration ellipses draw focus to the distribution of a group with the presence (red) or absence (green) of mesh exposure. Centroids of the concentration ellipses (large symbols) indicate the mean of each group. PC: principal component.



Predictive Analysis

Four supervised machine learning models incorporating all 13 cytokines were trained using 70% (42/60) of the available 60 observations (20 participants × 3 independent measurements); the remaining 30% (18/60) was used to test the models' accuracy when predicting the presence of mesh exposure through the vaginal wall. All 4 machine learning machines achieved at least 62% (26/42) training accuracy (Table 1). Artificial neural network achieved the highest prediction accuracy of 83% (15/18), while decision tree and Naïve Bayes both achieved a prediction accuracy of 61% (11/18). Naïve Bayes, decision tree, and artificial neural network excelled at correctly predicting patients with the presence of mesh exposure postsurgery at 89% (16/18). Artificial neural network was superior for correctly predicting patients who did not experience mesh exposure postsurgery (14/18, 78%).

XSL•FO
RenderX

Table 1. Summary of supervised learning model statistics. All 13 cytokines were used to predict the presence or absence of postsurgical mesh exposure through the vaginal wall; 70% (42/60) of observations were used for training, and 30% (18/60) of observations were used for testing.

Model	Training accuracy, n (%)	Prediction accuracy, n (%)	95% CI	Sensitivity, n (%)	Specificity, n (%)	Prediction, κ
Artificial neural network	33 (79)	15 (83)	0.586-0.964	14 (78)	16 (89)	0.667
Decision tree	27 (64)	11 (61)	0.57-0.827	6 (33)	16 (89)	0.222
Naïve Bayes	26 (62)	11 (61)	0.357-0.827	6 (33)	16 (89)	0.222
Logistic regression	31 (73)	9 (50)	0.260-0.740	10 (56)	8 (44)	0.000

Predictive Analysis Using Feature Selection

Additional models and predictive analyses explored whether a smaller set of cytokines could achieve similar predictive results. Feature selection using a random forest method identified a group of 7 cytokines capable of yielding effective predictive analysis: IL-1β, IL-8, IL-12p40, IL-12p70, TNF-α, IL-17A, and IL-6. Figure 4 illustrates that models exhibited variation among the importance of cytokines when implementing this more targeted group of cytokines. IL-1 and IL-8 are strongly represented in all models, while IL-6 is important in only Naïve Bayes.

Figure 4. A random forest algorithm was used to identify a group of 7 cytokines capable of yielding effective predictive analysis. The importance of each cytokine is evident in individual supervised learning models: ANN (A), DT (B), NB (C), and LR (D). IL-1α: interleukin-1 alpha; IL-1β: interleukin-1 beta; IL-2: interleukin-2; IL-4: interleukin-4; IL-6: interleukin-6; IL-8: interleukin-8; IL-10: interleukin-10; IL-12p40: interleukin-12p40; IL-12p70: interleukin-12p70; IL-17A: interleukin-17A; IFN-γ: interferon gamma; TNF-α: tumor necrosis factor-alpha; GM-CSF: granulocyte-macrophage colony-stimulating factor; PC: principal component; ANN: artificial neural network; DT: decision tree; NB: Naïve Bayes; LR: logistic regression.

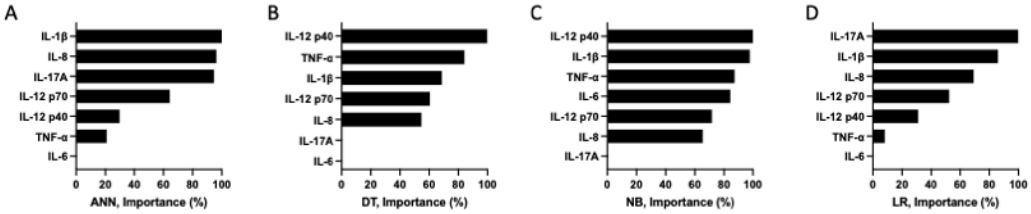


Table 2 illustrates that all models achieved at least 64% (27/42) training accuracy. The logistic regression model that used the 7 selected cytokines achieved a training accuracy of 81% (34/42), a prediction accuracy of 72% (13/18), a sensitivity of 67% (12/18), and a specificity of 78% (14/18), and thus outperformed compared to the logistic regression model that incorporated all of the cytokine data. Moreover, decision tree models achieved the same result when using the selected cytokines or when all cytokines were included. The prediction accuracy in Naïve Bayes and artificial neural network models executed with the 7 selected cytokines decreased by only 5% each compared to the same models that used all of the cytokine data.

Table 2. Summary of supervised learning model statistics. Feature selection via random forest was used to identify a group of 7 cytokines capable of yielding effective predictive analysis. The subset of cytokines was used to predict the presence or absence of postsurgical mesh exposure through the vaginal wall; 70% (42/60) of observations were used for training, and 30% (18/60) of observations were used for testing.

Model	Training accuracy, n (%)	Prediction accuracy, n (%)	95% CI	Sensitivity, n (%)	Specificity, n (%)	Prediction, κ
Artificial neural network	34 (81)	14 (78)	0.524-0.936	12 (67)	16 (89)	0.556
Decision tree	27 (64)	11 (61)	0.356-0.827	6 (33)	16 (89)	0.222
Naïve Bayes	30 (72)	10 (56)	0.308-0.785	6 (33)	14 (78)	0.111
Logistic regression	34 (81)	13 (72)	0.465-0.903	12 (67)	14 (78)	0.444

Discussion

Summary

Among patients with POP who undergo mesh implantation surgery, 17% of them experience mesh exposure through the vaginal wall [18]. This rate of surgical mesh complication is significant when compared to 0.035%-5.4% mesh-related erosions reported in other mesh-based surgeries [19-23], necessitating the development of a personalized decision support tool for patients with POP. This exploratory study demonstrates a novel and efficient approach to predicting postsurgical outcomes for mesh implantation using cytokine levels in patient blood following exposure to a biomaterial. Previous studies have often used patient demographic and medical data to train machine learning programs to create predictive outcomes for POP mesh surgeries [13]. In contrast, this study uses biological material to mimic an in vivo response, thus presenting a novel, noninvasive, personalized clinical decision tool. A systematic PCA approach identifies associations among cytokines that provide physiological insight. Supervised machine learning models developed in this study demonstrate that blood cytokine

XSL•FO
RenderX

measurements may be used as a predictive tool. In addition, the number of cytokine measurements needed may be reduced without compromising predictive capabilities, rendering this approach more applicable within a clinical setting.

Principal Findings and Comparison to Prior Work

The PCA analysis illustrated in Figure 1 reveals several significant associations among the cytokines. Several cytokines display positive correlations when comparing the 2 different stimuli: LPS and polypropylene mesh. However, IL-6 and IL-12p40 are negatively correlated between these 2 treatments. Thus, these 2 cytokines may explain the different inflammatory responses induced by LPS versus polypropylene mesh. When comparing blood samples incubated alone to those incubated in the presence of polypropylene mesh, only IL-1α exhibits a negative correlation, demonstrating that this proinflammatory mediator might be specifically affected by the mesh stimulus. Furthermore, 2 pairs of cytokines positively correlate (IL-10 and IL-4; IL-12p70 and IFN-γ), indicating that one of the cytokines in each pair could be eliminated to reduce the number of cytokines tested in a clinical setting.

The cytokines that contribute most to each PC segregate into proinflammatory and anti-inflammatory cytokines (Figure 2). When cytokine data from patient blood incubated with mesh were analyzed using PCA, cytokines IFN-γ, IL-12p70, and IL-2 were the largest contributors to the variance explained in PC 1. These markers are identified as proinflammatory agents [24-26], which suggests that proinflammatory cytokines may heavily influence PC 1. In contrast, cytokines IL-10, IL-4, and IL-6 were the largest contributors to the variance explained in PC 2. IL-4 and IL-10 are prominent anti-inflammatory cytokines [25], suggesting that anti-inflammatory cytokines heavily influence PC 2. IL-6, previously thought to have proinflammatory function only, is recently recognized as potentially having both proinflammatory and anti-inflammatory roles in COVID-19 [27] and diabetes [28].

When juxtaposing the biplot of polypropylene-stimulated cytokine observations (Figure 1B) with that of mesh exposure outcome (Figure 3), it can be extrapolated that proinflammatory cytokines IL-12p40, IL-1α, and TNF-α are positioned in the region of the biplot that uniquely corresponds to surgical outcomes involving the presence of mesh exposure through the vaginal wall. Such juxtaposition suggests that IL-12p40, IL-1α, and TNF-α may be associated with the presence of postsurgical mesh exposure. These observations may inspire potential therapeutic strategies that could improve postsurgical outcome. For example, the surgical mesh could be designed to modulate these key proinflammatory cytokines. In this way, while supporting the pelvic structure, the mesh could simultaneously function in controlling the cytokine response to minimize biomaterial rejection.

Table 1 describes the accuracy of supervised machine learning models and demonstrates that cytokines can exhibit predictive capabilities. Previous studies have performed predictive analysis for POP using risk factors derived from patient medical history [13]. However, such data can be incomplete and inaccurate [29]. This study demonstrates the utility that measured responses of biological samples can also have in developing robust predictive models. Chen et al [12] similarly used blood cytokine levels in a machine learning study to predict severe AKI after cardiac surgery. Their study concluded that a logistic regression model was the most effective in discovering the cytokine associations in severe AKI. In this study, the prediction accuracy for all 4 models exceeded 60%, with the artificial neural network model demonstrating the best overall performance, predicting POP surgical outcomes with 83% (15/18) accuracy when trained with all 13 cytokines. This predictive capability is similar to that reported for prediction derived from patient medical history [4], despite this study comprising a significantly smaller patient group. Considering the small population size, these results represent relatively high prediction accuracy for health care data.

When creating models trained with a subgroup of 7 cytokines (Table 2), selected using a random forest method, the artificial neural network model maintained the greatest effectiveness with respect to sensitivity, specificity, and prediction accuracy. Moreover, the group of selected cytokines outperformed the larger group of cytokines in the logistic regression model and achieved the same results in the decision tree model. The Naïve Bayes and artificial neural network prediction accuracy dropped only 5% when using the subgroup of cytokines, thus demonstrating the resiliency of these models. These results demonstrate that predictive capabilities are retained with fewer cytokines, which would enhance clinical feasibility by reducing the cost and time associated with this clinical decision tool.

Limitations and Future Directions

This study implemented rigorous research methods to identify physiological relationships among cytokine markers and developed robust machine learning models to predict mesh exposure; yet, some limitations should be noted. First, because this is a pilot study, the sample size is limited to 20 participants within a single hospital system. Nevertheless, this limited sample size predicted 83% (15/18) accuracy, a level that compares favorably with another predictive model study by Chu et al [9] that achieved a prediction accuracy of 71% in a study population size of 467. Thus, the results of this pilot study indicate the utility of this approach and the merit of future studies. Future study will provide validation with a larger population of participants from multiple hospitals. Additionally, the 10 participants in each group were not matched regarding variables. To minimize confounders, patients with POP recurrence or taking medication that would alter inflammatory response were excluded and the age ranges and average age at the time of surgery within each group were similar. Future studies with a larger population, however, will benefit from matching participants with respect to these and other potentially confounding variables. Nonetheless, the results of this pilot study highlight the importance of inflammatory markers in the prediction of this postsurgical condition.

Conclusions

While this preliminary study is limited to a sample size of just 20 participants, this novel approach to using cytokine response to predict POP surgical outcomes has successfully distinguished important cytokines and their correlations. Moreover, these relationships point toward prospective therapies that could

XSL•FO
RenderX

promote better surgical outcomes. Supervised learning models also demonstrate a high level of accuracy, specificity, and sensitivity, even when a smaller group of cytokine data is used. This result suggests that blood cytokine analysis might be feasibly used in a clinical setting to predict POP surgical outcomes. Further study with a larger patient population will be needed to confirm the utility of this method. If successful at a larger scale, this approach has the potential to change perspectives in which surgeons would recommend and proceed with POP repair surgeries and to prevent undesired outcomes of mesh-related surgeries in patients.

Acknowledgments

This work was supported by a Transformative Research Grant from Greenville Health System Health Sciences Center to MM, TW, RH, and RG.

Data Availability

Data are openly available in the GitHub public repository [30].

Conflicts of Interest

None declared.

References

1. Balsamo R, Illiano E, Zucchi A, Natale F, Carbone A, Sio MD, et al. Sacrocolpopexy with polyvinylidene fluoride mesh for pelvic organ prolapse: mid term comparative outcomes with polypropylene mesh. Eur J Obstet Gynecol Reprod Biol 2018 Jan;220:74-78. [doi: 10.1016/j.ejogrb.2017.11.018] [Medline: 29175131]
2. American College of Obstetricians and Gynecologists. Pelvic organ prolapse: ACOG practice bulletin, number 214. Obstet Gynecol 2019;134(5):e126-e142 [FREE Full text] [doi: 10.1097/AOG.0000000000003519] [Medline: 31651832]
3. Chughtai B, Barber MD, Mao J, Forde JC, Normand ST, Sedrakyan A. Association between the amount of vaginal mesh used with mesh erosions and repeated surgery after repairing pelvic organ prolapse and stress urinary incontinence. JAMA Surg 2017 Mar 01;152(3):257-263. [doi: 10.1001/jamasurg.2016.4200] [Medline: 27902825]
4. Reid FM, Elders A, Breeman S, Freeman RM, PROSPECT study group. How common are complications following polypropylene mesh, biological xenograft and native tissue surgery for pelvic organ prolapse? A secondary analysis from the PROSPECT trial. BJOG 2021 Dec;128(13):2180-2189 [FREE Full text] [doi: 10.1111/1471-0528.16897] [Medline: 34473896]
5. Mangir N, Roman S, Chapple CR, MacNeil S. Complications related to use of mesh implants in surgical treatment of stress urinary incontinence and pelvic organ prolapse: infection or inflammation? World J Urol 2020 Jan;38(1):73-80 [FREE Full text] [doi: 10.1007/s00345-019-02679-w] [Medline: 30759272]
6. Jones K. Chapter 9 - Fibrotic response to biomaterialsall associated sequence of fibrosis. In: Badylak SF, editor. Host Response to Biomaterials: The Impact of Host Response on Biomaterial Selection. Oxford, UK: Academic Press; 2015:189-237.
7. DiEgidio P, Friedman HI, Gourdie RG, Riley AE, Yost MJ, Goodwin RL. Biomedical implant capsule formation: lessons learned and the road ahead. Ann Plast Surg 2014 Oct;73(4):451-460. [doi: 10.1097/SAP.0000000000000287] [Medline: 25046663]
8. Raji CG, Vinod Chandra SS. Long-Term Forecasting the Survival in Liver Transplantation Using Multilayer Perceptron Networks. IEEE Trans Syst Man Cybern, Syst 2017 Aug;47(8):2318-2329 [FREE Full text] [doi: 10.1109/tsmc.2017.2661996]
9. Chu CS, Lee NP, Adeoye J, Thomson P, Choi SW. Machine learning and treatment outcome prediction for oral cancer. J Oral Pathol Med 2020 Nov;49(10):977-985. [doi: 10.1111/jop.13089] [Medline: 32740951]
10. Tseng PY, Chen YT, Wang CH, Chiu KM, Peng YS, Hsu SP, et al. Prediction of the development of acute kidney injury following cardiac surgery by machine learning. Crit Care 2020;24(1):478 [FREE Full text] [doi: 10.1186/s13054-020-03179-9] [Medline: 32736589]
11. Akbari H, Rathore S, Bakas S, Nasrallah MP, Shukla G, Mamourian E, et al. Histopathology-validated machine learning radiographic biomarker for noninvasive discrimination between true progression and pseudo-progression in glioblastoma. Cancer 2020;126(11):2625-2636 [FREE Full text] [doi: 10.1002/cncr.32790] [Medline: 32129893]
12. Chen Z, Chen L, Yao G, Yang W, Yang K, Xiong C. Novel blood cytokine-based model for predicting severe acute kidney injury and poor outcomes after cardiac surgery. J Am Heart Assoc 2020 Nov 17;9(22):e018004 [FREE Full text] [doi: 10.1161/JAHA.120.018004] [Medline: 33131359]
13. Jelovsek JE, Chagin K, Lukacz ES, Nolen TL, Shepherd JP, Barber MD, NICHD Pelvic Floor Disorders Network. Models for predicting recurrence, complications, and health status in women after pelvic organ prolapse Surgery. Obstet Gynecol 2018 Aug;132(2):298-309 [FREE Full text] [doi: 10.1097/AOG.0000000000002750] [Medline: 29995735]

XSL•FO
RenderX

14. Lê S, Josse J, Husson F. FactoMineR: a package for multivariate analysis. J Stat Soft 2008;25(1):1-18. [doi: 10.18637/jss.v025.i01]
15. Venables WN, Ripley BD. Modern Applied Statistics with S. 4th ed. New York, USA: Springer; 2002.
16. Joanes DN, Gill CA. Comparing measures of sample skewness and kurtosis. J Royal Statistical Soc D 1998 Mar;47(1):183-189 [FREE Full text] [doi: 10.1111/1467-9884.00122]
17. Kuhn M. Building predictive models in R using the caret package. J Stat Softw 2008;28(5):1-26. [doi: 10.18637/jss.v028.i05]
18. Ubertazzi EP, Soderini HFE, Saavedra Sanchez AJM, Fonseca Guzman C, Paván LI. Long-term outcomes of transvaginal mesh (TVM) In patients with pelvic organ prolapse: a 5-year follow-up. Eur J Obstet Gynecol Reprod Biol 2018 Jun;225:90-94. [doi: 10.1016/j.ejogrb.2018.03.060] [Medline: 29680466]
19. Spiro C, Quarmby N, Gananadha S. Mesh-related complications in paraoesophageal repair: a systematic review. Surg Endosc 2020;34(10):4257-4280 [FREE Full text] [doi: 10.1007/s00464-020-07723-0] [Medline: 32556700]
20. Sánchez-Pernaute A, Pérez-Aguirre ME, Jiménez AP, Campos AR, Muñoz A, Torres A. Intraluminal mesh erosion after prosthetic hiatoplasty: incidence, management, and outcomes. Dis Esophagus 2019 Jul 01;32(7):1-5. [doi: 10.1093/dote/doy131] [Medline: 30791045]
21. Koelbl H, Stoerer S, Seliger G, Wolters M. Transurethral penetration of a tension-free vaginal tape. BJOG 2001 Jul;108(7):763-765. [doi: 10.1111/j.1471-0528.2001.00188.x] [Medline: 11467707]
22. Novara G, Galfano A, Boscolo-Berto R, Secco S, Cavalleri S, Ficarra V, et al. Complication rates of tension-free midurethral slings in the treatment of female stress urinary incontinence: a systematic review and meta-analysis of randomized controlled trials comparing tension-free midurethral tapes to other surgical procedures and different devices. Eur Urol 2008 Feb;53(2):288-308. [doi: 10.1016/j.eururo.2007.10.073] [Medline: 18031923]
23. Taylor D, Barton E. In vitro characterisation of the erosion of soft tissues by surgical mesh. J Mech Behav Biomed Mater 2020 Jan;101:103420. [doi: 10.1016/j.jmbbm.2019.103420] [Medline: 31518946]
24. Xu X, Wang G, Ai L, Shi J, Zhang J, Chen YX. Melatonin suppresses TLR9-triggered proinflammatory cytokine production in macrophages by inhibiting ERK1/2 and AKT activation. Sci Rep 2018;8(1):15579 [FREE Full text] [doi: 10.1038/s41598-018-34011-8] [Medline: 30349079]
25. Miranda TS, Heluy SL, Cruz DF, da Silva HDP, Feres M, Figueiredo LC, et al. The ratios of pro-inflammatory to anti-inflammatory cytokines in the serum of chronic periodontitis patients with and without type 2 diabetes and/or smoking habit. Clin Oral Investig 2019;23(2):641-650 [FREE Full text] [doi: 10.1007/s00784-018-2471-5] [Medline: 29737428]
26. Zhao Y, Zhang T, Guo X, Wong CK, Chen X, Chan YL, et al. Successful implantation is associated with a transient increase in serum pro-inflammatory cytokine profile followed by a switch to anti-inflammatory cytokine profile prior to confirmation of pregnancy. Fertil Steril 2021;115(4):1044-1053 [FREE Full text] [doi: 10.1016/j.fertnstert.2020.10.031] [Medline: 33272613]
27. Scherger S, Henao-Martínez A, Franco-Paredes C, Shapiro L. Rethinking interleukin-6 blockade for treatment of COVID-19. Med Hypotheses 2020;144:110053 [FREE Full text] [doi: 10.1016/j.mehy.2020.110053] [Medline: 32758889]
28. Akbari M, Hassan-Zadeh V. IL-6 signalling pathways and the development of type 2 diabetes. Inflammopharmacology 2018 Jun;26(3):685-698. [doi: 10.1007/s10787-018-0458-0] [Medline: 29508109]
29. Dunn S, Lanes A, Sprague AE, Fell DB, Weiss D, Reszel J, et al. Data accuracy in the Ontario birth registry: a chart re-abstraction study. BMC Health Serv Res 2019;19(1):1001 [FREE Full text] [doi: 10.1186/s12913-019-4825-3] [Medline: 31881960]
30. BiomaterialStudyCytokine. GitHib. URL: https://github.com/tomhjlim/BiomaterialStudyCytokine.git [accessed 2023-05-12]

Abbreviations

AKI: acute kidney injury
GM-CSF: granulocyte-macrophage colony-stimulating factor
IFN: interferon
IL: interleukin
IRB: institutional review board
LPS: lipopolysaccharide
PC: principal component
PCA: principal component analysis
POP: pelvic organ prolapse
TNF: tumor necrosis factor

XSL•FO
RenderX

Edited by T Leung; submitted 21.06.22; peer-reviewed by S Azadnajafabad, K Bera, SAH Aqajari; comments to author 25.02.23; revised version received 13.04.23; accepted 30.04.23; published 31.05.23.

Please cite as:
Waugh ML, Boltin N, Wolf L, Goodwin J, Parker P, Horner R, Hermes M, Wheeler T, Goodwin R, Moss M
Prediction of Pelvic Organ Prolapse Postsurgical Outcome Using Biomaterial-Induced Blood Cytokine Levels: Machine Learning Approach
JMIR Perioper Med 2023;6:e40402
URL: https://periop.jmir.org/2023/1/e40402
doi:10.2196/40402
PMID:37256676

©Mihyun Lim Waugh, Nicholas Boltin, Lauren Wolf, Jane Goodwin, Patti Parker, Ronnie Horner, Matthew Hermes, Thomas Wheeler, Richard Goodwin, Melissa Moss. Originally published in JMIR Perioperative Medicine (http://periop.jmir.org), 31.05.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Perioperative Medicine, is properly cited. The complete bibliographic information, a link to the original publication on http://periop.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Early Warning Scores to Support Continuous Wireless Vital Sign Monitoring for Complication Prediction in Patients on Surgical Wards: Retrospective Observational Study

Mathilde C van Rossum[1,2], PhD; Robin E M Bekhuis[3,4], MSc; Ying Wang[1,4], PhD; Johannes H Hegeman[3], MD, PhD; Ellis C Folbert[3], PhD; Miriam M R Vollenbroek-Hutten[1], PhD; Cornelis J Kalkman[5], MD, PhD; Ewout A Kouwenhoven[3*], MD, PhD; Hermie J Hermens[1*], PhD

[1]Department of Biomedical Signals and Systems, University of Twente, Enschede, Netherlands
[2]Department of Cardiovascular and Respiratory Physiology, University of Twente, Enschede, Netherlands
[3]Department of Surgery, Hospital Group Twente, Almelo, Netherlands
[4]Hospital Group Twente Academy, Hospital Group Twente, Almelo, Netherlands
[5]Department of Anesthesiology, University Medical Center Utrecht, Utrecht, Netherlands
[*]these authors contributed equally

Corresponding Author:
Mathilde C van Rossum, PhD
Department of Cardiovascular and Respiratory Physiology
University of Twente
PO box 217
Enschede, 7500 AE
Netherlands
Phone: 31 53 489 4152
Email: m.c.vanrossum@utwente.nl

Abstract

Background: Wireless vital sign sensors are increasingly being used to monitor patients on surgical wards. Although early warning scores (EWSs) are the current standard for the identification of patient deterioration in a ward setting, their usefulness for continuous monitoring is unknown.

Objective: This study aimed to explore the usability and predictive value of high-rate EWSs obtained from continuous vital sign recordings for early identification of postoperative complications and compares the performance of a sensor-based EWS alarm system with manual intermittent EWS measurements and threshold alarms applied to individual vital sign recordings (single-parameter alarms).

Methods: Continuous vital sign measurements (heart rate, respiratory rate, blood oxygen saturation, and axillary temperature) collected with wireless sensors in patients on surgical wards were used for retrospective simulation of EWSs (sensor EWSs) for different time windows (1-240 min), adopting criteria similar to EWSs based on manual vital signs measurements (nurse EWSs). Hourly sensor EWS measurements were compared between patients with (event group: 14/46, 30%) and without (control group: 32/46, 70%) postoperative complications. In addition, alarms were simulated for the sensor EWSs using a range of alarm thresholds (1-9) and compared with alarms based on nurse EWSs and single-parameter alarms. Alarm performance was evaluated using the sensitivity to predict complications within 24 hours, daily alarm rate, and false discovery rate (FDR).

Results: The hourly sensor EWSs of the event group (median 3.4, IQR 3.1-4.1) was significantly higher ($P<.004$) compared with the control group (median 2.8, IQR 2.4-3.2). The alarm sensitivity of the hourly sensor EWSs was the highest (80%-67%) for thresholds of 3 to 5, which was associated with alarm rates of 2 (FDR=85%) to 1.2 (FDR=83%) alarms per patient per day respectively. The sensitivity of sensor EWS–based alarms was higher than that of nurse EWS–based alarms (maximum=40%) but lower than that of single-parameter alarms (87%) for all thresholds. In contrast, the (false) alarm rates of sensor EWS–based alarms were higher than that of nurse EWS–based alarms (maximum=0.6 alarm/patient/d; FDR=80%) but lower than that of single-parameter alarms (2 alarms/patient/d; FDR=84%) for most thresholds. Alarm rates for sensor EWSs increased for shorter time windows, reaching 70 alarms per patient per day when calculated every minute.

XSL•FO
RenderX

Conclusions: EWSs obtained using wireless vital sign sensors may contribute to the early recognition of postoperative complications in a ward setting, with higher alarm sensitivity compared with manual EWS measurements. Although hourly sensor EWSs provide fewer alarms compared with single-parameter alarms, high false alarm rates can be expected when calculated over shorter time spans. Further studies are recommended to optimize care escalation criteria for continuous monitoring of vital signs in a ward setting and to evaluate the effects on patient outcomes.

(JMIR Perioper Med 2023;6:e44483) doi:10.2196/44483

KEYWORDS

early warning scores; vital signs; telemedicine; physiological monitoring; clinical alarms; postoperative complications; perioperative nursing

Introduction

Background

Surgical patients are at risk of developing postoperative complications, which may progress to life-threatening illnesses and seriously affect patient outcomes if not promptly detected and correctly treated [1]. Most postoperative complications occur in the first week after surgery and are typically present in a ward setting [2-5]. Therefore, adequate patient monitoring in surgical wards is crucial for identifying the early signs of complications [6].

In hospital wards, patient monitoring typically consists of routine vital sign checks performed by nurses every 6 to 8 hours, complemented by subjective evaluation of the patient status during nursing activities [7,8]. In addition, early warning scores (EWSs) are widely used to evaluate the risk of patient deterioration. EWSs are typically calculated by assigning points to a measured set of vital signs, where the sum of the points reflects the EWSs. EWSs are often implemented as part of rapid response systems, where they are used to trigger clinical actions or escalation of care when exceeding a predefined threshold [5,9]. However, vital sign checks and corresponding EWSs are often incomplete or not performed on time, particularly during the night or when the protocol mandates more frequent measurements in high-risk patients [10,11]. Together with the intermittent measurement frequency, this may lead to unnoticed or delayed detection of patient deterioration.

In recent years, wireless sensors that enable mobile vital sign monitoring have been introduced. These sensors facilitate automated, less obtrusive, and continuous patient monitoring in a ward setting [12,13]. Although there is still little evidence regarding the clinical effects of continuous monitoring in this setting, various studies have suggested that continuous monitoring can aid early identification of clinical deterioration and may provide opportunities to improve outcomes in patients with complications [4,14]. However, the interpretation of the large amount of data that are generated by the sensors is still a major challenge because vital sign measurements fluctuate largely during the day and are influenced by movement and many patient-related or environmental factors [15]. Moreover, continuous manual data observation is hampered by restricted staffing levels in a ward setting and inconsistent assessment of abnormalities by caregivers [16]. Therefore, to promote an adequate and timely response to patient deterioration, automatic methods to support and identify vital sign abnormalities related to potential complications are desired.

Currently, wireless monitoring is often implemented in combination with traditional alarm strategies, where an alert is sent automatically as soon as the measurement of one of the vital signs exceeds a preset upper or lower threshold [12,17]. Although this single-parameter alarm strategy is standard in high-care units such as intensive care units, these alarms are sensitive to various disturbances and easily lead to excessive false alarm rates. As the nurse:patient ratio is lower in a ward setting, this alarm burden is a serious concern for nurse workload and could lead to alarm fatigue, thereby potentially threatening patient safety [7]. Furthermore, single-parameter abnormality detection does not align with the current use of EWSs for risk prediction in patients on surgical wards, which relies on multiple parameters.

Objective

Accordingly, it may be of interest to use EWSs instead of single-parameter alarms to detect potential abnormalities in continuously monitored patients on surgical wards, as supported by recently developed remote monitoring systems [12]. However, there is still little evidence regarding whether EWSs derived from mobile vital sign measurements can actually support the identification of deterioration in patients on surgical wards and be useful as alarm systems for continuous remote monitoring. Therefore, this exploratory study aimed to gain insight into the potential sensitivity of a sensor-based, high-rate EWS for predicting postoperative complications and to evaluate the expected daily alarm rate in comparison with single-parameter alarm criteria as well as manual EWS measurements for different alarm settings. Furthermore, the potential predictive value of “nurse worry” and patient-reported deterioration was explored to investigate the possible benefits of systematically collecting subjective data in monitoring routines.

Methods

Study Design

The study had an observational retrospective design.

Ethics Approval

This study was approved by The Medical Research Ethics Committee Twente (MoViSign study; NL65885.044.18).

XSL•FO
RenderX

Informed Consent

All included participants provided written informed consent to participate in the study and to use their data for research purposes.

Population

The study included patients (aged ≥18 years) undergoing elective esophageal or gastric resection (ie, upper gastrointestinal [GI] group) and older patients (aged ≥70 years) undergoing acute surgery for a hip fracture (ie, hip fracture group), as these groups are known to have relatively high rates of complications during ward stay. Patients were recruited preoperatively or after postoperative admission to the surgical ward (GI unit or center of geriatric traumatology). Patients were excluded if they had implanted electronic medical devices, had known allergies to materials used in wearable sensors, had suspected delirium or cognitive impairment, or were unable to decide upon study participation. Recruitment was performed during office hours only, and patients who were expected to leave the hospital within 24 hours after possible recruitment were not eligible for study participation.

Measurements

The patients received standard ward care including routine nursing observations. Nurses were instructed to perform manual measurements of vital signs and calculate a corresponding Modified Early Warning Score (MEWS) [18] at least 3 times a day. The MEWS was used to indicate the risk of patient deterioration (≤2=low risk, 3-4=intermediate risk, and ≥5=high risk), where cutoff values of 3 and 5 were used as response triggers to call a physician and the rapid response team for further patient investigation, respectively. In addition to routine care, vital signs were recorded every minute using a wearable patient monitoring system. Heart rate (HR) and respiratory rate (RR) were measured using the chest-worn LifeTouch sensor (Isansys Lifecare Ltd). Axillary temperature (AT) was recorded with the LifeTemp sensor (Isansys Lifecare Ltd), and blood oxygen saturation (SpO_2) was measured with the 3150 WristOx finger pulse oximeter (Nonin Medical Inc). All the measurements were sent to a bedside patient gateway using an encrypted Bluetooth connection. Sensor recordings were started after the patient was admitted to the surgical ward and informed consent was obtained and continued until hospital discharge (or premature patient dropout). During the recording period, the researchers checked and maintained the technical functionality of the sensor recordings at least once a day on weekdays. Medical professionals and patients were blinded to the sensor measurements to prevent bias, but nurses were instructed to temporarily detach the sensors for showering, diagnostic imaging, or surgical interventions or reinterventions.

To explore the added value of routine collection of subjective information from nurses and patients, nurses were instructed to fill in a paper checklist during every nurse shift to register the possible presence and reasons for nurse worry using an adapted version of the Dutch Early Nurse Worry Indicator Score, as specified in Multimedia Appendix 1 [19]. Furthermore, patients were asked to fill in a daily diary during their ward stay to indicate the presence of symptoms, how they were currently feeling (0 points=very poor; 5 points=very good), and how they felt compared with the previous day (0 points=much impaired; 5 points=much improved) on a 5-point Likert scale. If needed, researchers helped or encouraged the patients to fill in the diary when they visited the patients. The nurse worry checklist and patient diary were used only for research purposes. All collected measurements were deidentified for further analysis.

Events

An event was defined as a postoperative complication that was diagnosed according to local standards and that was treated during ward stay or within 7 days after hospital discharge. Events were identified retrospectively from patient records and reviewed by a surgeon (EAK or HJH) to ensure correct interpretation. As there is no way to ascertain exactly when a complication started, we recorded the start of targeted treatment as the onset time of the complication. Similarly, the end time of the complication was defined as the timing of the last therapeutic action, for example, the last medication gift. If no information regarding the last therapeutic action was available or if therapy was continued after hospital discharge (or premature patient dropout) or in case of palliative treatment, it was assumed that the complication lasted until the end of the sensor recording period. Complications classified as Clavien-Dindo class [20] of II or higher, for which treatment was started within the measurement period (ie, the period of sensor measurements during ward stay), were eligible and included for further analysis. In patients with multiple complications, eligible events were only included if treatment was started at least 24 hours after any previously included complication. Furthermore, complications were only included if at least 4 hours of HR and RR sensor measurements were available in the 24 hours window before the onset time of the event after data preprocessing (see the *Data Preprocessing* section). Patients with eligible complications were enrolled in the event group, whereas patients with an uncomplicated postoperative trajectory served as the control group. Patients with only ineligible complications were excluded from further analyses. To evaluate the risk of selection bias, the baseline characteristics were compared between the included and excluded patient groups.

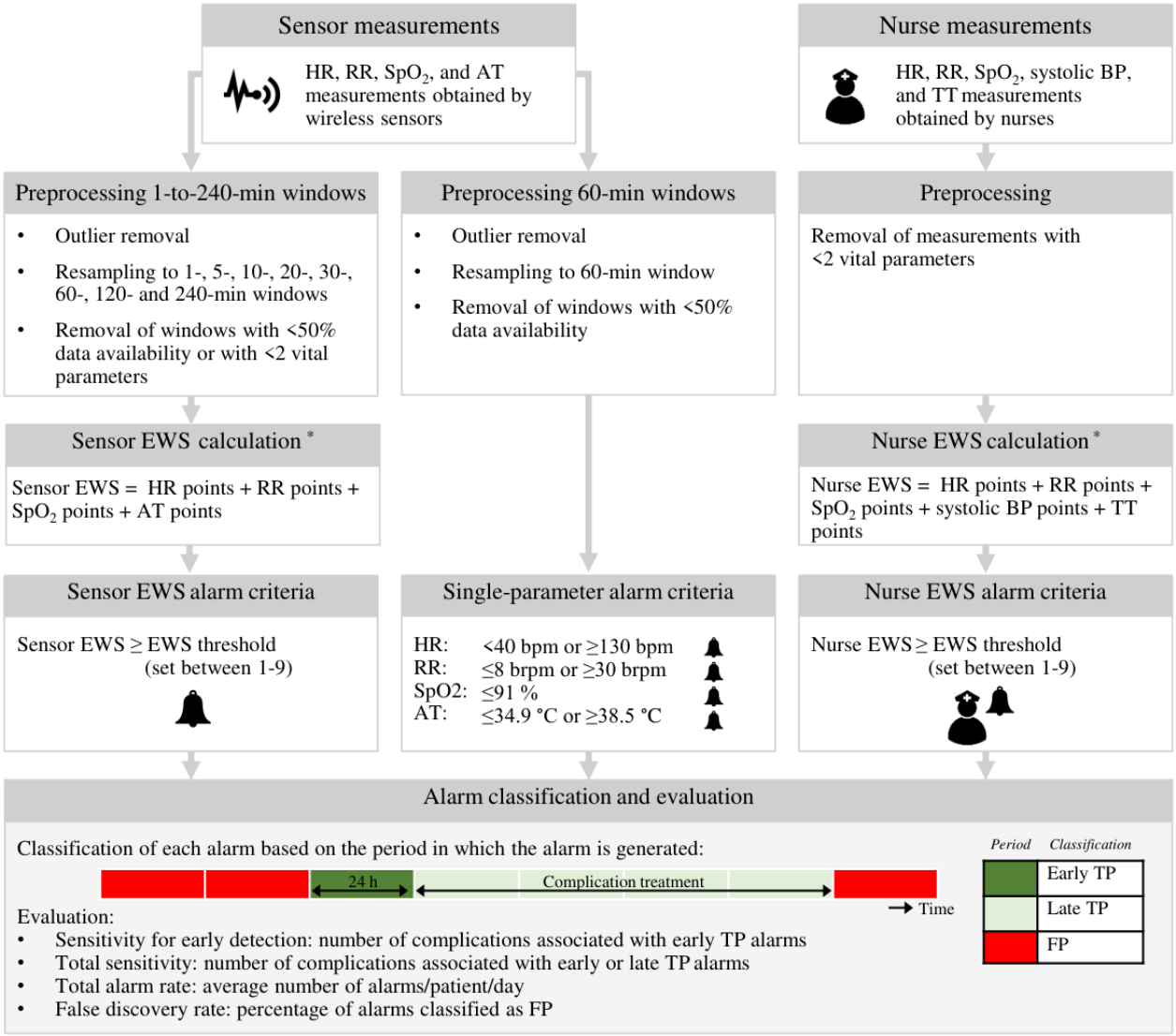
EWS and Alarm Simulation

Overview

Sensor and nurse measurements were used to simulate corresponding EWSs. In addition, alarms were respectively simulated using the EWSs based on sensor vital signs measurements (sensor EWSs), EWSs based on manual vital signs measurements (nurse EWSs), and single vital sign recordings, with the aim of evaluating and comparing the effects of using these measurements as response triggers or active alarms. Figure 1 provides an overview of the alarm simulation and evaluation, which are further explained in the following subsections.

XSL•FO
RenderX

Figure 1. Overview of the calculation of the early warning score (EWS), alarm simulation, and alarm evaluation in sensor and nurse measurements. AT: axillary temperature; BP: blood pressure; bpm: beats per minute; brpm: breaths per minute; FP: false-positive; HR: heart rate; nurse EWS: early warning score based on manual vital signs measurements; RR: respiratory rate; sensor EWS: early warning score based on sensor vital signs measurements; SpO_2: blood oxygen saturation; TP: true-positive; TT: tympanic temperature. *EWS calculation criteria are described in Table 1.



XSL•FO
RenderX

Table 1. Criteria used to calculate early warning scores (EWSs) based on sensor vital signs measurements (sensor EWSs) and early warning scores based on manual vital signs measurements (nurse EWSs).

Variable	Inclusion of vital sign		Points						
	Sensor EWSs	Nurse EWSs	3	2	1	0	1	2	3
RR[a] (brpm[b])	✓[c]	✓		≤8		9-14	15-20	21-29	≥30
HR[d] (bpm[e])	✓	✓		≤40	41-50	51-100	101-110	111-129	≥130
SpO_2[f] (%)	✓	✓[g]	≤91	92-93	94-95	≥96			
Systolic BP[h] (mm Hg)		✓	≤70	71-80	81-100	101-200		≥201	
AT[i] or TT[j] (°C)[k]	✓	✓		≤34.9		35-38.4		≥38.5	

[a]RR: respiratory rate.
[b]brpm: breaths per minute.
[c]Variables included in the sensor EWSs and nurse EWSs.
[d]HR: heart rate.
[e]bpm: beats per minute.
[f]SpO_2: blood oxygen saturation.
[g]No standard protocol; therefore, only included if available.
[h]BP: blood pressure.
[i]AT: axillary temperature.
[j]TT: tympanic temperature.
[k]AT used for sensor EWSs and TT used for nurse EWSs.

Data Preprocessing

Data preprocessing and analysis were performed using MATLAB (version R2021b; The Math Works Inc). Sensor recordings were preprocessed by removing implausible extreme values (HR >200 or <30 beats/min [bpm], RR <5 or >50 breaths/min [brpm], AT <30 °C or >50 °C, and SpO_2 <70% or >100%), most likely caused by artifacts. As no standard monitoring frequency has currently been established for monitoring patients on surgical wards, sensor measurements were analyzed repeatedly for different time windows. Accordingly, the minute-sampled vital sign recordings were resampled by averaging the signal values in successive windows of 5, 10, 20, 30, 60, 120, and 240 minutes. For each vital parameter, windows wherein data were missing for >50% of the time were disregarded.

EWS Simulation

The original and down-sampled sensor measurements as well as all vital sign measurements obtained by nurses during the sensor recording period were used for the retrospective simulation of the corresponding EWSs. For this purpose, points were assigned to each vital parameter according to the MEWS criteria [18] (Table 1). As SpO_2 is not included in the MEWS, the SpO_2 criteria were obtained from the UK National Early Warning Score criteria [21].

The sensor EWSs were calculated as the sum of the points assigned to each of the vital parameters measured by the sensor system, resulting in sensor EWSs between 0 and 11. Similarly, the vital sign measurements obtained by nurses were used to calculate the nurse EWSs, ranging between 0 and 14. The sensor EWSs and nurse EWSs were only calculated for measurements in which at least 2 vital sign values were available.

Alarm Simulation

To evaluate and compare the effect of using the sensor or nurse measurements as an active notification system, “alarms” were simulated based on the sensor EWS measurements (“sensor EWS–based alarms”) and based on all available nurse EWS measurements (“nurse EWS–based alarms”). An alarm was defined as an occurrence where the EWS increased to a higher level compared with a previous measurement, thereby exceeding a preset threshold (“EWS threshold”). When the measurements exceeded the EWS threshold from the beginning of the measurement period, the first sample was considered as the first alarm. EWS alarms were simulated repeatedly for EWS thresholds of 1 to 9, aiming to evaluate the alarms for a wide range of thresholds. However, as MEWS cutoff values of 3 and 5 were used in this study’s hospital as response triggers, these thresholds were selected as EWS threshold in the primary analysis. For the sensor EWSs, the primary investigation was based on 1-hour time windows, but alarms were also simulated for all other time windows (1-240 min).

In addition to the alarms based on EWSs, sensor measurements were used to simulate single-parameter alarms, similar to traditional physiological alarm systems. A single-parameter alarm was simulated when the vital sign measurement exceeded the predefined normal range. Upper and lower thresholds were defined by the outer limits of the EWS criteria, that is, HR ≤40 or ≥130 bpm, RR ≤8 or ≥30 brpm, SpO_2 ≤91%, and AT ≤34.9 °C or ≥38.5 °C. Single-parameter alarms were simulated for 1-hour time windows, and situations where multiple single-parameter alarms were present in the same window were counted as 1 alarm.

XSL•FO
RenderX

Alarm Classification

All the simulated alarms were classified based on the timing of the alarms, as illustrated in Figure 1. Alarms that were simulated ≤24 hours before the onset of the included events were classified as "early true-positive" (TP) alarms, thereby reflecting alarms that could theoretically promote early identification of events. Alarms that presented during the treatment period of the included events were classified as "late true-positive" alarms. Alarms that presented >24 hours before onset or after the end of the treatment period of included events or that presented in the control group were classified as false-positive (FP) alarms, that is, false alarms.

Analysis

Availability and Agreement of Vital Signs Measurements

The availability of sensor measurements, that is, data completeness, was calculated as the percentage of the total recording time where vital sign measurements were available after preprocessing in the total study population. The availability of nurse measurements was verified by the number of measurements available per 24 hours. The agreement between nurse and sensor measurements was explored for each vital parameter using Bland-Altman analysis for repeated measures [22] by comparing each available nurse measurement with the corresponding average value of the available preprocessed sensor measurements in the 5-minute window [23] before the nurse measurement.

Group Comparison

All the measurements (sensor recordings, sensor EWSs, nurse EWSs, nurse worry checklist, and patient diary) were compared between the event and control groups to explore their relationships with the development of complications. For all available hourly sensor EWS and nurse EWS measurements obtained during the sensor recording period, the average EWSs and the percentage of measurements where the EWSs were ≥3 (intermediate to high risk) and ≥5 (high risk) was investigated for each group. For the sensor EWSs, the number of EWS points assigned to the hourly sensor measurement windows was evaluated to investigate the contribution of the different vital parameters to the sensor EWSs. Furthermore, the percentage of hourly sensor measurements, that is, single-parameter recordings, that exceeded the upper or lower thresholds defined for the single-parameter alarms were assessed. Finally, the prevalence of worry expressed by nurses and the prevalence of (very) poor current status and (severely) impaired status indicated by patients during the study period were evaluated to explore the potential diagnostic value of this subjective information. All group differences were statistically compared using the Mann-Whitney *U* test. In addition, a sensitivity analysis was performed to verify the impact of patient exclusion on group differences by separately evaluating group differences after adding excluded patients to the control group.

Alarm Evaluation

The performance of the sensor EWS–based, nurse EWS–based, and single-parameter alarms was investigated using 2 sensitivity rates: the total alarm rate (TAR) and the false discovery rate (FDR) [24]. The sensitivity for early detection of adverse events was calculated as the percentage of events associated with early TP alarms. Similarly, the total sensitivity for event detection was defined as the percentage of events for which early TP and late TP alarms were observed. The TAR was defined as the average daily number of alarms per patient and the FDR as the percentage of alarms classified as FP. The sensitivity for early detection, total sensitivity, TAR, and FDR were compared between the sensor EWS–based, nurse EWS–based, and single-parameter alarms.

Results

Population

A total of 60 patients were included in the study, of whom 33 (55%) were patients undergoing elective esophageal or gastric resection and 27 (45%) were older patients with a hip fracture. The baseline characteristics of the participants are presented in Table 2. On average, sensor recordings were started 48 hours after the onset of surgery. Out of 60 patients, the period of vital sign recording was completed by 46 (77%) patients, whereas measurements were stopped before hospital discharge in 14 (23%) patients due to withdrawal from study participation (9/14, 64%), patient transfer (2/14, 14%), organizational reasons (2/14, 14%), and transition to palliative care (1/14, 7%). No temperature sensor was placed in 2 patients, and pulse oximeter recordings were not started in 2 other patients due to unavailable sensors or patient refusal.

XSL•FO
RenderX

Table 2. Baseline characteristics of the study population.

Characteristic	Upper GI[a] group (n=33)	Hip fracture group (n=27)
Age (years), mean (SD)	64 (11)	82 (7)
Sex, n (%)		
Male	24 (73)	4 (15)
Female	9 (27)	23 (85)
American Society of Anesthesiologists classification, n (%)		
II	20 (61)	12 (44)
III	12 (36)	11 (41)
IV	1 (3)	3 (11)
Unknown	0 (0)	1 (4)
Comorbidities, n (%)		
Cardiovascular	19 (58)	19 (70)
Diabetes	3 (9)	7 (26)
GI	12 (36)	8 (30)
Neuropsychiatric	6 (18)	10 (37)
Pulmonary	6 (18)	8 (30)
Urogenital	9 (27)	11 (41)
Other categories[b]	6 (18)	11 (41)
ICU[c] readmission, n (%)	3 (9)	0 (0)
In-hospital mortality, n (%)	0 (0)	0 (0)
Out-of-hospital mortality[d], n (%)	0 (0)	1 (4)
Hospital readmission[d], n (%)	0 (0)	0 (0)

[a]GI: gastrointestinal.

[b]Includes endocrine, infective, neuromuscular, or thrombosis-related comorbidities.

[c]ICU: intensive care unit.

[d]Within 7 days of hospital discharge.

Events

During the measurement period or the 7-day follow-up period, a total of 40 complications were observed in 28 patients. Of all observed complications (n=40), 25 (63%) events were excluded (Clavien-Dindo class I: 10/25, 40%; within 24 hours from the previous event: 2/25, 8%; no or insufficient sensor data availability: 13/25, 52%), resulting in the exclusion of 14 patients from the analysis. Out of 40 events, 15 (38%) were included in the analysis (Clavien-Dindo class II: 10/15, 67%; III: 2/15, 13%; IV: 1/15, 7%; V: 1/15, 7%). The included events were present in a group of 14 patients (upper GI group: n=9, 64%; hip fracture group: n=5, 36%), who were selected as the event group. The control group consisted of 32 patients (upper GI group: n=17, 53%; hip fracture group, n=15, 47%) in whom no events were observed. We found no statistical differences in the baseline characteristics between the event and control groups or between the included and excluded patient groups.

Availability and Agreement of Vital Signs Measurements

Sensor recordings in the total study population had a median duration of 120 (IQR 93-163) hours. After preprocessing, the median data availability of HR and RR was 84% (IQR 74%-94%) of the measurement time and that of AT was 97% (IQR 84%-100%). SpO_2 values were missing more frequently, with a median availability of 46% (IQR 40%-60%). The median number of vital sign observations registered by nurses was 3 (IQR 2-4) per day. On average, HR was available in 96%, RR in 48%, SpO_2 in 92%, systolic blood pressure in 95%, and tympanic temperature in 93% of nurse observations.

Bland-Altman plots for all available nurse and sensor data pairs are presented in Multimedia Appendix 2. The sensor measurements of HR and RR were higher than the nurse measurements, with a mean difference of 3 bpm and 9.4 brpm, respectively, where the largest differences between measurements were particularly seen for higher measurement values. In contrast, the SpO_2 (mean absolute difference 2.1%) and temperature measurements (mean difference 0.9 °C)

XSL•FO
RenderX

recorded by the sensor were lower than the nurse measurements, where the largest deviations were observed in the lower ranges. For all parameters, the limits of agreement were relatively wide compared with the observed range of values.

Group Comparison

Figure 2 shows the hourly sensor EWSs over time for the event and control groups, visualized using EWS thresholds of 3 and 5. High sensor EWSs were often scattered over time, and no uniform sensor EWS pattern was observed in the days or hours preceding the complication treatment. According to the average number of points that were assigned to the hourly vital sign measurements (Figure 3), RR and SpO_2 received 2 or 3 points in at least half of the available measurements for both the event and control groups, thereby contributing most to the sensor EWSs.

Table 3 describes the differences in the measurements between the event and control groups. The average hourly sensor EWSs of the included patients with complications were significantly higher (*P*=.004) than those of the patients with uncomplicated postoperative trajectories. In addition, the sensor EWSs in the event group reached scores of ≥5 more often compared with the control group (*P*<.001). For nurse EWSs, the number of measurements, as well as the average nurse EWSs and the percentage of observations where sensor EWSs of ≥3 or ≥5, were significantly higher for the event group (*P*=.02, *P*=.009, *P*<.001, and *P*<.001, respectively). Furthermore, nurses expressed worry more often for the patients in the event group (*P*=.02). Indicators for nurse worry that were reported in both groups are specified in Multimedia Appendix 1. In the event group, most reported indicators of worry included "Change in breathing," "Subjective nurse observation," and "Patient indicates." The patient diary results did not differ significantly between the groups (*P*=.35 and *P*=.37). However, patients in the event group seemed to report a (very) poor and a (much) impaired status more often, although these patients filled in the daily diary less frequently. According to the sensitivity analysis, adding the excluded patient group to the control group did not change any of the results, as all *P* values remained within the same levels of significance (ie, *P*<.001, *P*<.05, or *P*>.05).

In the 24 hours before the onset of the included events, the average sensor EWSs had a median value of 3.4 (IQR 2.5-4.0), which was significantly different from the median value of 2.0 (IQR 1.1-2.7) of the average nurse EWSs in this period (*P*=.006) and from the median value of 2.8 (IQR 2.4-3.2) of the average sensor EWSs observed across the total measurement period in the control group (*P*=.04).

Figure 4 shows a case example of a patient undergoing esophagectomy diagnosed on postoperative day 8 with anastomotic leakage (Clavien-Dindo class IV) using a computed tomography scan. Treatment included the placement of an esophageal stent on day 9 (defined as event onset) and antibiotic therapy, followed by intensive care unit readmission on day 11 due to progressive hemodynamic instability. The sensor recordings contained missing data periods for all the vital parameters. Furthermore, the sensor measurements differed from the nurse measurements, particularly in terms of RR and temperature. The sensor and nurse measurements showed a similar gradually increasing pattern in HR in the days before event diagnosis and treatment. However, the sensor measurements indicated an increasing, although extreme, trend in RR before diagnosis, whereas nurse RR measurements increased only after the stent placement. Correspondingly, the sensor EWSs reached values ≥5 more frequently from day 7, whereas the nurse EWSs reached values of ≥5 only on day 10. In addition to abnormal vital sign measurements, nurse worry as well as impaired or poor self-reported patient status were observed not only in the 24-hour period before stent placement but also earlier in the postoperative trajectory.

XSL•FO
RenderX

Figure 2. Early warning score based on sensor vital signs measurements (sensor EWS) of all patients enrolled in the event and control group, calculated using 1-hour windows and presented according to cutoff values of ≥3 (orange) and ≥5 (red). Complication timing (onset of treatment) is annotated by white diamond markers.

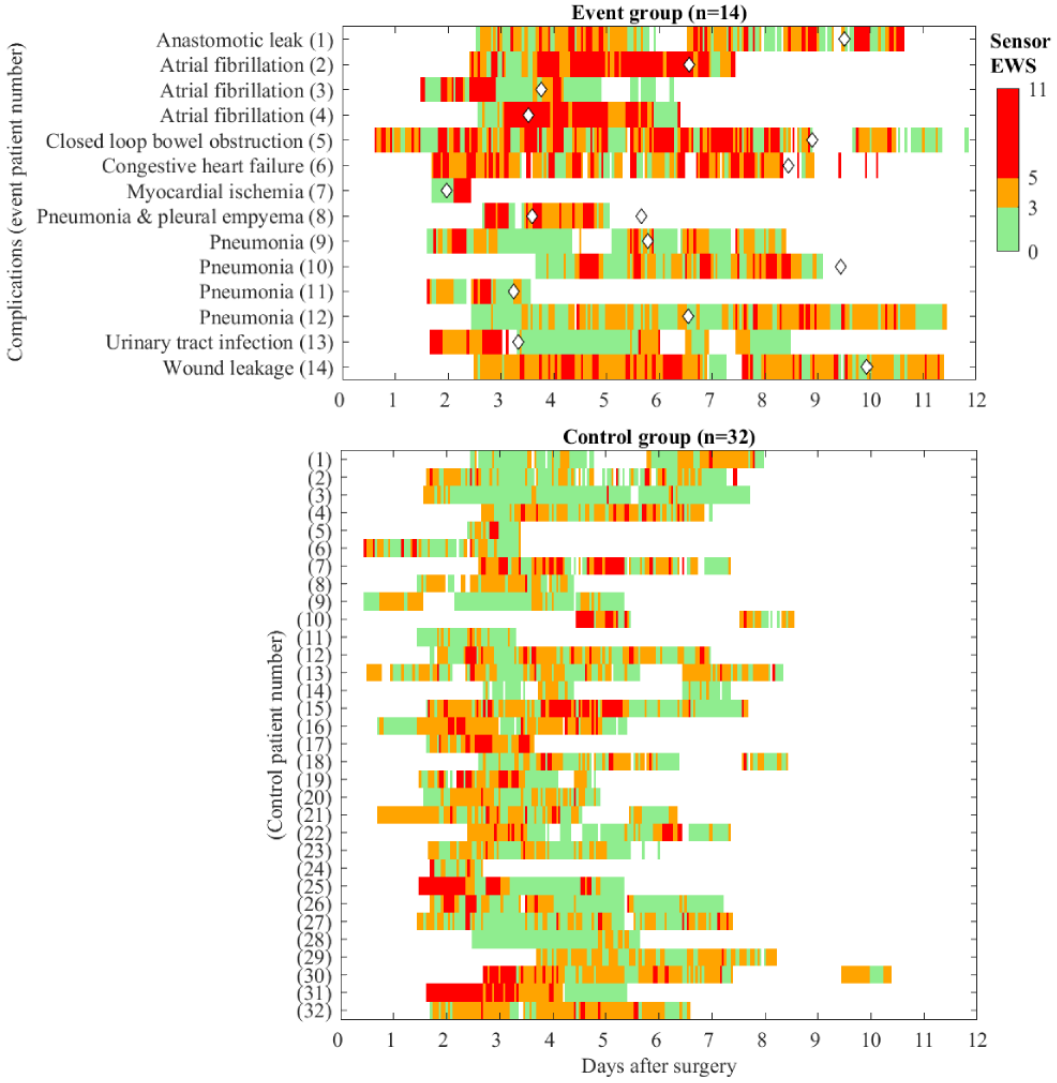
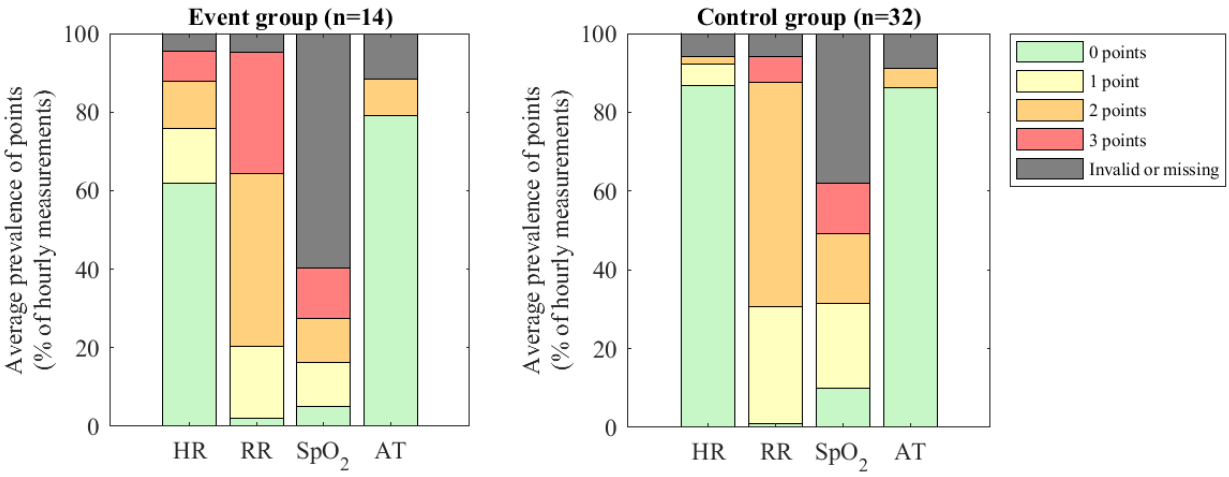


Figure 3. Number of points that were assigned to the hourly sensor vital sign measurements in patients with and without postoperative complications following the criteria used to calculate early warning scores (specified in Table 1). The bar stacks reflect the average percentage of 1-hour windows in which 0, 1, 2, or 3 points were assigned to the corresponding vital parameter. AT: axillary temperature; HR: heart rate; RR: respiratory rate; SpO_2: blood oxygen saturation.



XSL•FO
RenderX

Table 3. Measurement results in the event and control group.

Result	Event group (n=14), median (IQR)	Control group (n=32), mean (IQR)	*P* value
Sensor measurements			
Measurement period[a] (h)	186 (92-260)	119 (106-142)	.10
Hourly sensor EWSs[b]			
Sensor EWS availability (% of the measurement period)	82 (66-98)	85 (68-93)	.94
Average sensor EWSs	3.5 (3.1-4.1)	2.8 (2.4-3.2)	.004
Sensor EWSs ≥3 (% of available hourly sensor measurements)	62 (44-83)	49 (41-66)	.08
Sensor EWSs ≥5 (% of available hourly sensor measurements)	25 (18-40)	8.1 (2.4-17)	<.001
Nurse EWSs[c]			
Nurse EWSs availability (number of measurements/24 h)	3.4 (2.3-4.2)	2.7 (1.9-3.1)	.02
Average nurse EWSs	1.7 (0.9-2.3)	0.9 (0.4-1.3)	.009
Nurse EWSs ≥3 (% of available measurements)	22 (7.7-34)	0 (0-12)	<.001
Nurse EWSs ≥5 (% of available measurements)	4.3 (0-6.9)	0 (0-0)	<.001
Hourly single-parameter recordings			
1 parameter out of normal range[d] (% measurement period)	26 (19-41)	18 (13-25)	.03
Multiple parameters out of normal range[d] (% measurement period)	3.0 (0.0-13)	0 (0-0.8)	<.001
Nurse worry checklist			
Nurse worry checklist availability (number of checklists filled/24 h)	1.2 (0.8-1.9)	1.2 (0.9-1.6)	.93
Nurse worry expressed (% of available nurse worry checklists)	25 (0-33)	0 (0-17)	.02
Patient diary			
Patient diary availability (number of diary forms filled/24 h)	0.5 (0-1)	0.8 (0-1)	.37
Patient expressed (very) poor and (much) impaired status (% of available diary forms)	14 (0-36)	0 (0-33)	.35

[a]Period of sensor measurements performed during ward stay.

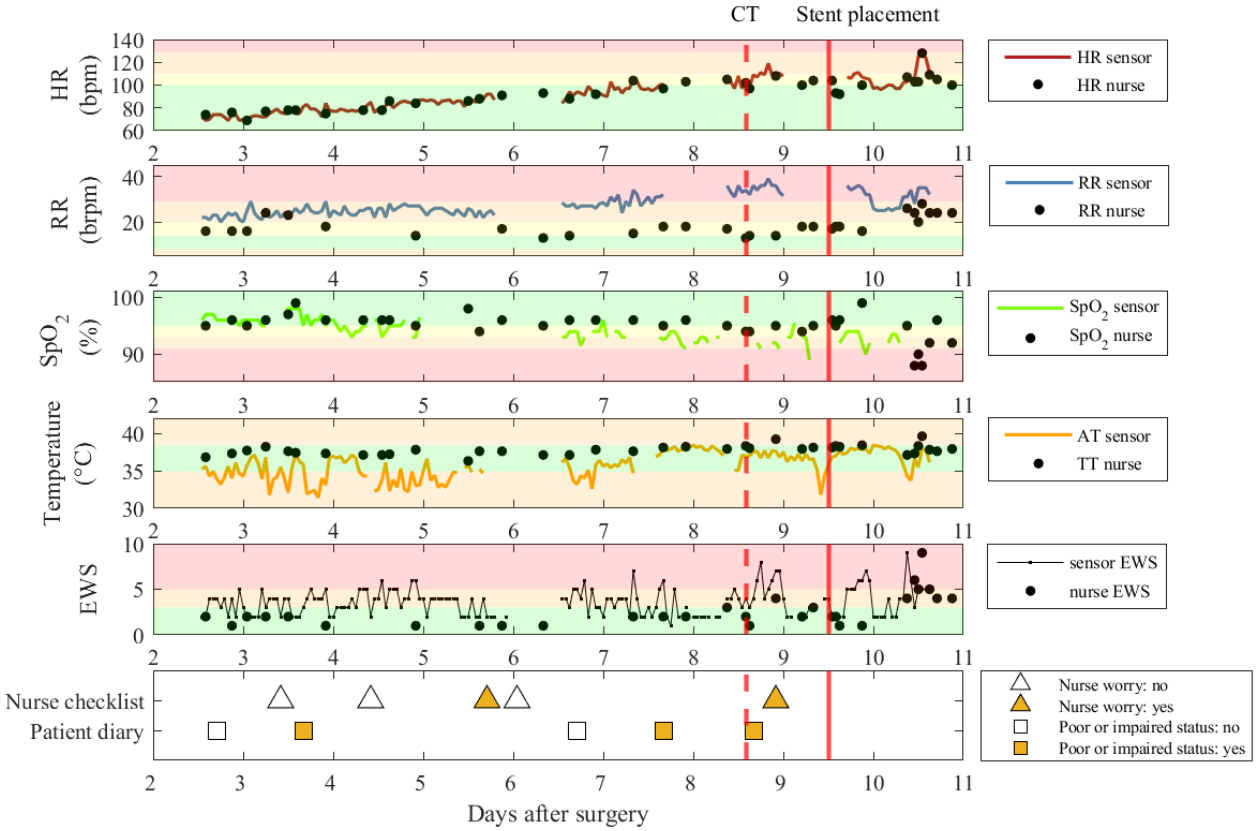
[b]Sensor EWS: early warning score based on sensor vital sign measurements.

[c]Nurse EWS: early warning score based on manual vital sign measurements.

[d]Out-of-normal range conforms to single-parameter alarm criteria.

XSL•FO
RenderX

Figure 4. Case example of a patient with anastomotic leakage. AT: axillary temperature; bpm: beats per minute; brpm: breaths per minute; CT: computed tomography; EWS: early warning score; HR: heart rate; nurse EWS: EWS based on manual vital signs measurements; RR: respiratory rate; SpO_2; blood oxygen saturation; sensor EWS: EWS based on sensor vital signs measurements; TT: tympanic temperature. The sensor vital sign measurements and sensor EWS measurements are calculated using 1-hour windows.



Alarm Evaluation

Sensor EWS-Based Alarms

Figure 5 displays the performance metrics of sensor EWSs alarms simulated in 1-hour windows using EWS thresholds of 1 to 9. The sensitivity rates were the highest for an EWS threshold of 3, where alarms were observed 24 hours before the onset of included events in 12 out of 15 events (sensitivity for early detection=80%) and during the treatment period for all events (total sensitivity=100%). TAR followed a similar pattern as the sensitivity rates and reached the highest level for a threshold of 3, resulting in a maximum of 2 alarms per patient per day, of which 85% was classified as FP alarms. An EWS threshold of 5 was associated with a TAR of 1.2 alarms per patient per day, FDR of 83%, sensitivity for early detection of 67%, and total sensitivity of 93%.

Figure 6 shows the performance of sensor EWSs alarms for window lengths of 1 to 240 minutes, illustrated for an EWS threshold of 5 as an example. The sensitivity rates and FDR decreased for longer windows. TAR strongly declined with increasing window length, ranging from 31 alarms per patient per day for window lengths of 1 minute to 0.8 alarms per patient per day for window lengths of 120 minutes. Similar trends were observed for all other thresholds, although the sensitivity decreased more rapidly with increasing window length for the other thresholds. In addition, FDR increased with the window length for thresholds of <3. The highest TAR levels were observed for a threshold of 3, reaching up to 70 alarms per patient per day for a 1-minute window length.

XSL•FO
RenderX

Figure 5. Sensitivity, total alarm rate (TAR), and alarm classification of early warning score based on sensor vital signs measurements (sensor EWS)–based alarms, calculated using 1-hour windows and early warning score (EWS) thresholds of 1 to 9. FP: false-positive; TP: true-positive.

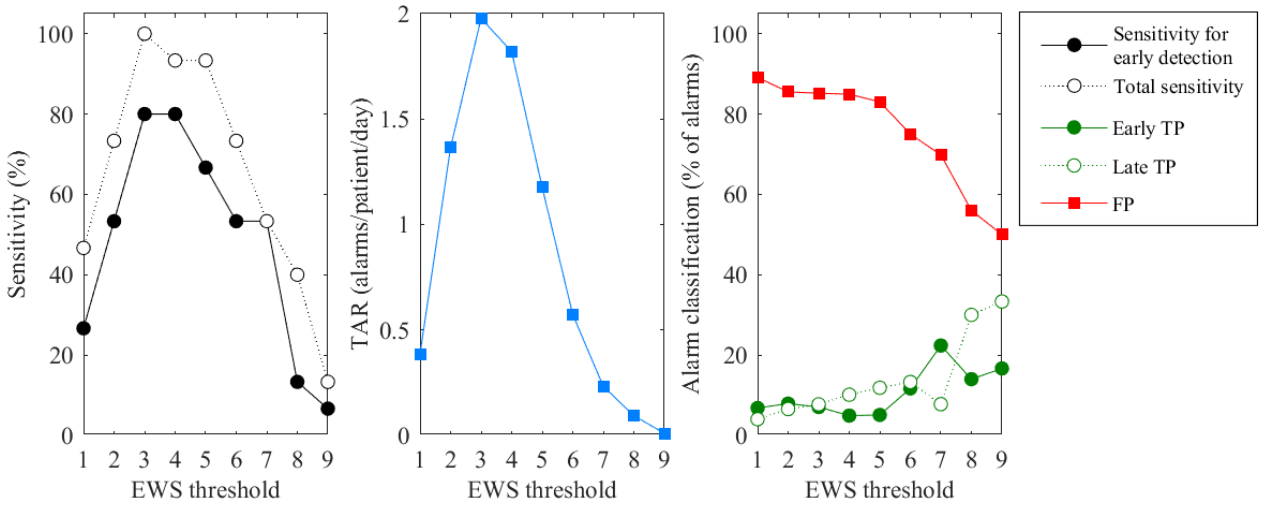
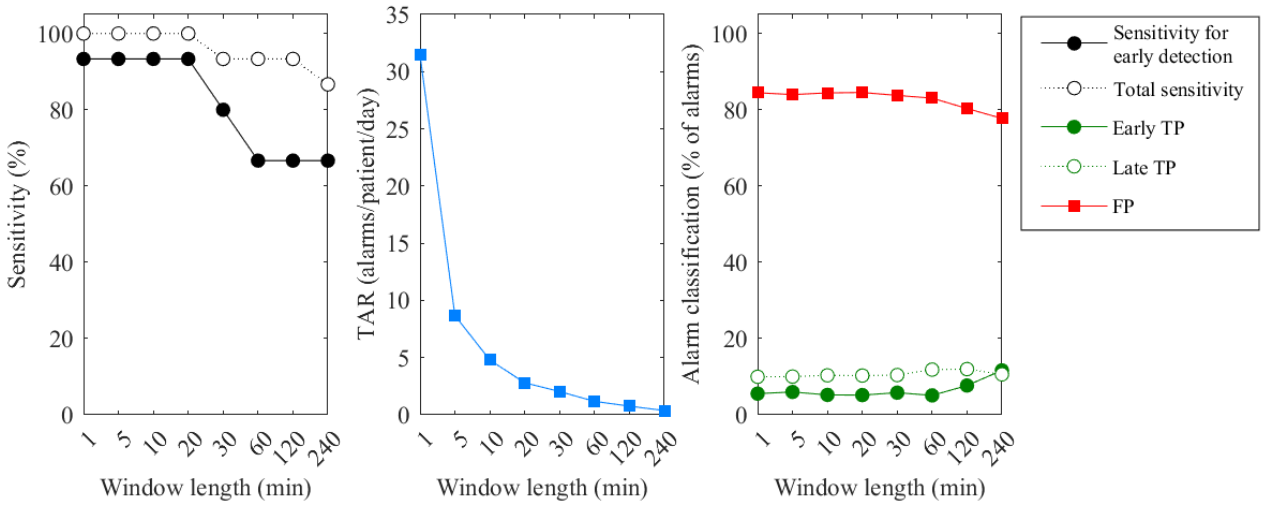


Figure 6. Sensitivity, total alarm rate (TAR), and alarm classification of early warning score based on sensor vital signs measurements (sensor EWS)–based alarms, calculated using 1- to 240-minute windows and an early warning score (EWS) threshold of 5. FP: false-positive; TP: true-positive.

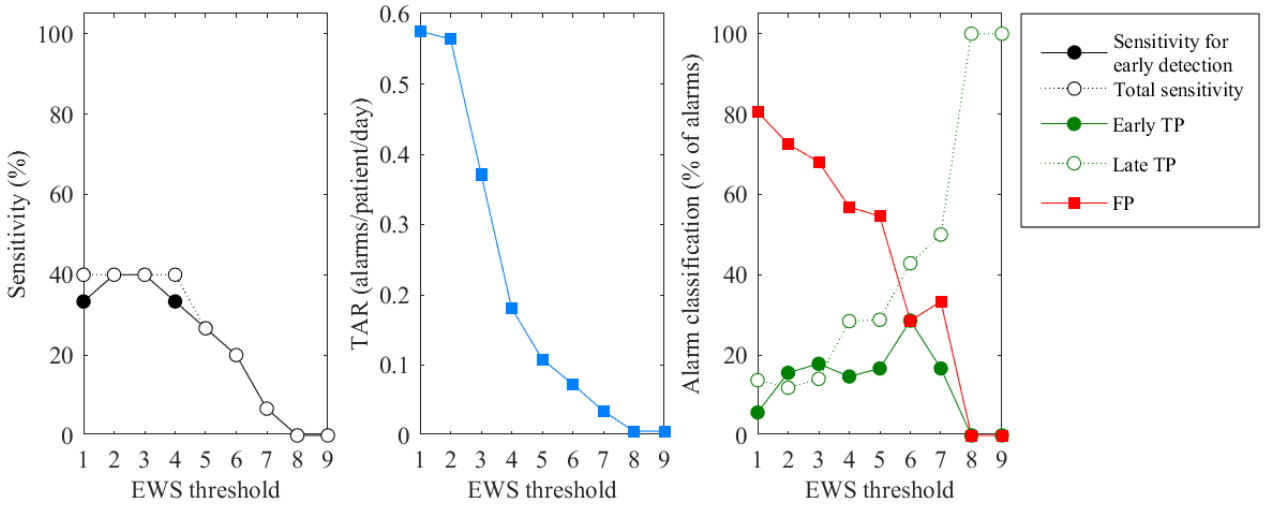


Nurse EWS-Based Alarms

Figure 7 displays the performance metrics of nurse EWS–based alarms. The sensitivity for early detection and total sensitivity were the same for most threshold values, reaching a maximum of 40%. The TAR and FDR were the highest (0.6 alarms/patient/d and 80%, respectively) for a threshold of 1 but decreased for higher threshold values. An EWS threshold of 5 was associated with a sensitivity for early detection and total sensitivity of 27%, TAR of 0.1 alarms per patient per day, and FDR of 54%. Compared with the sensor EWSs, not only the sensitivity rates but also the TAR and FDR of the nurse EWSs were lower for all thresholds >1.

XSL•FO
RenderX

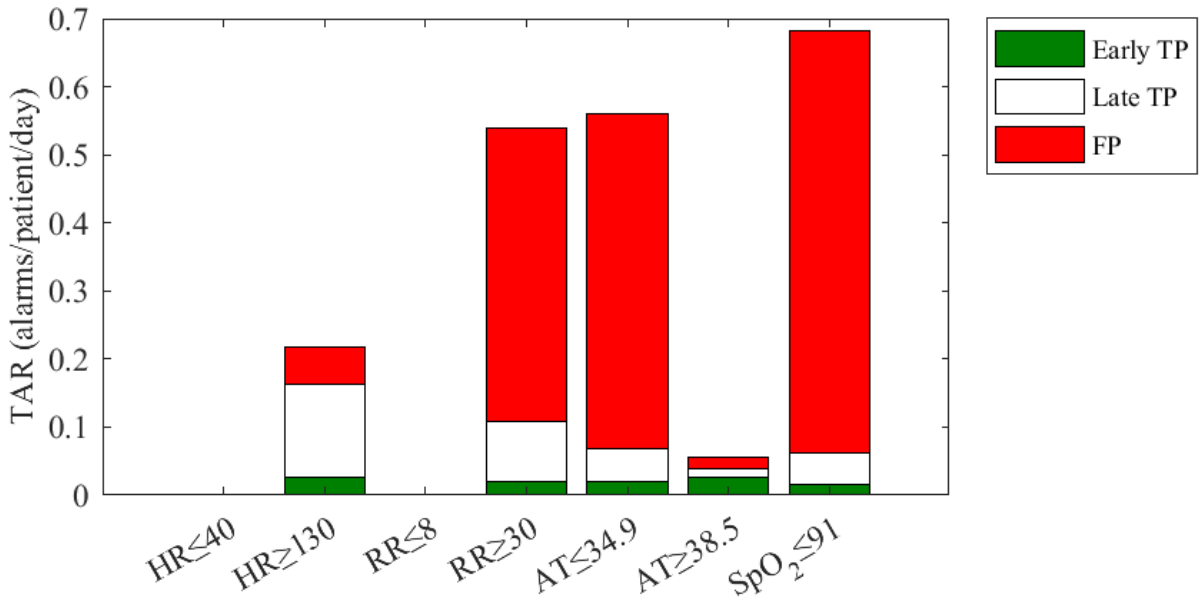
Figure 7. Sensitivity, total alarm rate (TAR), and alarm classification of early warning score based on manual vital signs measurements (nurse EWS)–based alarms, calculated using early warning score (EWS) thresholds of 1 to 9. FP: false-positive; TP: true-positive.



Single-Parameter Alarms

Figure 8 shows the TAR and classification of the single-parameter alarms. In total, the single-parameter alarms resulted in a TAR of 2 alarms per patient per day. The FDR of all single-parameter alarms was 83.6%, and 4.7% of the alarms were classified as early TP and 11.7% as late TP. In 13 of the 15 events, at least 1 single-parameter alarm was observed in the 24-hour period before the onset of the event, resulting in a sensitivity for early detection of 87%. Total sensitivity was 100%. Most alarms (>0.5 alarms/patient/d) were observed for low SpO_2, low AT, and high RR, of which a large part (≥80%) was classified as FP alarms. In contrast, high HR and high AT alarms were associated with lower alarm rates and lower FDR (≤30%). No single-parameter alarms were observed for low HR or low RR.

Figure 8. Total alarm rate (TAR) of single-parameter alarms calculated using 1-hour windows. AT: axillary temperature; FP: false-positive; HR: heart rate; RR: respiratory rate; SpO_2: blood oxygen saturation; TP: true-positive.



Discussion

Principal Findings

This study investigated the potential usability of high-rate EWSs obtained from wireless vital sign data for early identification of complications in patients on surgical wards and explored the performance of a sensor EWS–based alarm system in comparison with single-parameter alarms and manual EWS measurements by nurses for various alarm settings. The EWSs based on hourly sensor recordings were significantly higher in patients with complications than in patients with an uncomplicated postoperative trajectory. Furthermore, the sensitivity to predict adverse events within 24 hours was higher for sensor EWS–based alarms compared with nurse EWS–based alarms. Therefore, EWSs obtained from sensor measurements might contribute to early awareness or confirmation of patient deterioration in current ward routines. However, high EWSs

XSL•FO
RenderX

were often recurrent in patients with and without events and could therefore lead to high false alarm rates when used as real-time alarm systems. Although hourly EWS measurements resulted in fewer alarms compared with single-parameter alarms, the number of false EWS alarms increased when using shorter time windows and varied between alarm thresholds, highlighting the importance of careful selection of alarm settings.

Comparison With Prior Work

Prediction of Complications

Traditionally, vital signs are used to monitor patients after surgery, as vital signs change 4 to 24 hours before adverse events [25]. The additional use of EWSs supports systematic assessment and responses to patient deterioration by nurses [5,26]. Coherently, it can be expected that frequent monitoring of EWSs can contribute to improved and early recognition of complications. This expectation is supported by the current case example and by other studies where abnormal vital signs and corresponding (partial) EWSs were observed more evidently, more often, or earlier in sensor recordings than in nurse measurements [23,27]. Similarly, sensor EWS–based alarms and single-parameter alarms based on hourly sensor measurements provide a higher sensitivity for predicting events within 24 hours compared with nurse EWS–based alarms. RR contributed the most to high sensor EWSs, which agrees with previous studies reporting RR as an important predictor of deterioration [5,7,28]. Furthermore, SpO_2 abnormalities were often observed, which is in line with the high prevalence of hypoxemic (micro) events found in similar observational studies [27,29].

Compared with hourly sensor EWS–based alarms, we found higher sensitivity rates for single-parameter alarms, although the differences were small for some EWS thresholds. This observation is in line with a study by Rothman et al [30] who reported that only 32% of life-threatening adverse events were preceded by abnormalities in multiple components of the EWS, whereas 27% showed only a single abnormal criterion. In contrast, Rothman et al [30] also reported that 41% of the adverse events were not preceded by abnormal vital signs at all. This raises the question of the relatively high sensitivity of both the sensor EWS–based and single-parameter alarms. Accordingly, it is plausible that abnormal vital sign measurements were not caused by the complication development but were related to normal variations within patients or the postoperative status, as supported by the observation that high sensor EWSs were also present in the control group. Similar findings were reported by Itelman et al [31], who reported that national EWSs based on 5-minute wireless vital sign recordings obtained in patients on general wards provided warnings for deterioration events in 67% of the cases on average 29 hours before detection by caregivers but also led to warnings in 78% of all patients who did not experience deterioration. Similarly, Haahr - Raunkjaer et al [29] reported that episodes of abnormal vital sign measurements were observed more often during the 24-hour period before events in patients admitted to the ward; however, the duration of abnormalities in continuous recordings did not differ between patients with and without serious adverse events.

Another possible explanation for the low specificity is that sensor-based respiration rates were typically much higher than nurse-based measurements. Whether this reflects the inaccuracy of nurse-based RR measurements [10] or RR overestimation by the sensor remains to be determined. In this context, one should be aware that the overestimation of RR falsely increases EWSs, which reduces the discriminative ability of systems. Together, these findings highlight the risk of event overdetection and highlight the need for further optimization of the combined warning criteria for wireless monitoring.

Alarms

The use of automatic alarm systems in monitoring routines can contribute to the early awareness and active responses of care professionals to potential deterioration [32]. However, excessive false alarm rates can lead to alarm fatigue, posing a serious risk to patient safety and a barrier to the adoption of continuous monitoring [33]. Therefore, a balanced ratio between alarm sensitivity, specificity, and alarm rate is crucial, which requires careful selection of alarm thresholds. In current practice, a variety of EWS thresholds are used, depending on the EWS variant and clinical settings. The calculation of the sensor EWSs was based on the MEWS criteria, for which studies recommended thresholds of 3 or 4 for monitoring patients on wards [34,35]. However, the results of this study indicate that the sensor EWSs resulted in the highest false alarm rates for EWS thresholds of 3 and 4. In addition, we found that sensor EWSs ranged between 2 and 4 and rarely dropped to values <2, which also explains the low alarm rates for thresholds of <3. Therefore, using a threshold of 5 may be considered to reduce the alarm burden in sensor-based alarm systems.

Alarm rates based on the hourly sensor EWSs were lower than those based on the hourly single-parameter alarms, which underlines the benefits of a multiple-parameter approach to improving specificity. However, it should be noted that the EWS alarm rates increased rapidly when using shorter time windows. Therefore, it might be appropriate to restrict automatic alarm generation based on EWSs to intervals of 1 hour or even longer in a ward setting and additionally present historical data trends for a more detailed evaluation of patient deterioration during nurse rounds. The use of summary measures for continuously measured vital signs is supported by van Goor et al [36], who reported that mean values calculated over multiple hour time frames can be helpful in supporting the identification of respiratory insufficiency. Although assessing data over longer time windows instead of every few minutes will limit alarm sensitivity, this approach would still be an improvement compared with intermittent nurse observations by allowing more frequent and less-static EWS measurements. Moreover, such an approach might be well accepted, as it protects nurses from alarm overload and promotes the use of the system as a support system without compromising the nurses' leading role in the decision-making process. For safe and effective implementation, it is crucial that caregivers are trained in interpreting the data and alarms and that high-risk patients requiring continuous surveillance are admitted to a high-care unit with a nurse:patient ratio that allows immediate response in case of acute deterioration [7,37].

XSL•FO
RenderX

Sensor Versus Manual Vital Signs Measurements

The use of wireless sensors provides new opportunities for monitoring vital signs in ward settings. By enabling continuous recording, wireless monitoring supports the comprehensive investigation of data abnormalities and trends, reducing the likelihood that signs of deterioration are missed due to intermittent or incomplete nurse measurements [10,27]. In addition, reducing the time needed for routine measurements can reduce nurses' workloads, leaving more time for patient observation and support. However, remote vital sign monitoring technology is still immature, and standards for the implementation of remote patient monitoring in ward routines are lacking. Furthermore, sensor measurements during patient mobilization still present practical challenges; for example, wireless connection issues and limited patient adherence [27,33] can lead to data loss. In our study, we encountered most measurement issues for SpO_2, where we found a relatively high rate of missing data related to detachment of the finger probe by patients, loss of Bluetooth connection, and short battery life, as also reported in another study using the same device [29]. Technological improvements can be achieved by developing new sensors that can derive accurate SpO_2 values from locations on the body that are less prone to (movement) artifacts and sensor dislodgment. Furthermore, adequate methods for correcting missing data periods in postanalysis and real-time alarm systems are desired.

Second, wearable sensor measurements are prone to artifacts caused by motion and poor skin contact, and clinical validation studies have shown variable accuracy of remote systems [38-40]. These sensor inaccuracies could explain the observed low agreement between the sensor and nurse data pairs, where the average bias was particularly large for RR and SpO_2 as also observed in other studies [23]. However, it must be noted that nurse observations also have variable accuracy, especially for RR [10]. In addition, discrepancies between the sensor and nurse measurements can be expected when measurements are performed using different sensing techniques, sensor locations, and under different circumstances; for example, in this study, sensor measurements recorded AT instead of tympanic temperature and were performed during daily activities, where vital sign measurements can be influenced by physical activity, whereas nurse measurements were performed during rest. Together, these factors underline that sensor measurements and corresponding EWSs are not a one-on-one replacement of measurements performed by nurses, where each approach has its benefits and limitations. These differences highlight the need for appropriate response trigger criteria for continuous monitoring, perhaps requiring different classifications of EWS points or the implementation of alternative detection algorithms.

Finally, when seeking continuous risk prediction methods or fully automatic EWS assessments, it should be considered that currently available wireless monitoring systems are restricted to measuring only a selection of vital parameters. For example, the level of consciousness is implemented in standard EWS or patient monitoring routines [9] but cannot be measured automatically using current sensor systems. Until reliable sensing and preprocessing techniques are available for all relevant parameters, it would be of interest to develop ways to optimally combine sensor and nurse measurements in care routines. Various commercially available systems for wireless or automatic patient monitoring already support or are used for the calculation of complete EWS every few hours [12,41], often after verification of valid measurement values and manual insertion of parameters that cannot yet be measured by the system, such as blood pressure or consciousness. Following this approach, Weenk et al [23] performed a study that included patients on wards, where an EWS was calculated every 30 minutes using 2 different wireless vital sign sensor systems and complementary nurse measurements. Higher EWSs were seen earlier in combined recordings than in nurse measurements alone, suggesting that integration of manual and automatic monitoring could contribute to better continuous patient monitoring and improve early identification of deterioration.

Nurse Worry and Patient-Reported Indicators

Subjective evaluation of patient status is an essential element of patient monitoring and clinical decision-making [7,37]. The presence of nurse worry has been described as a predictor of deterioration and is increasingly used as an individual calling criterion or as part of EWS variations [19,42]. The importance of nurse worry was confirmed by this study, where we found that a nurse worry checklist could support the identification of patients with events and may reveal early signs of deterioration, as illustrated by the case example. Accordingly, wireless monitoring should be accompanied by routine nurse evaluations, which could mitigate the risk of overreliance on technology [7]. To ensure consistent collection and valuation of subjective information that cannot be retrieved with sensors, it can be beneficial to embed structured bedside observation scores such as the Dutch Early Nurse Worry Indicator Score [19] in electronic patient files and warning criteria.

In addition to nurse worry, well-being indicators reported by patients or by their relatives provide important signs of deterioration [8], and the presence of patient concerns has been suggested as an additional care escalation criterion [43]. Correspondingly, we found that patient-reported deterioration was often reported as a reason for nurse worry in patients in the event group. However, signs of deterioration are not always recognized or expressed by patients or their relatives [43]. In this study, exploring a patient diary that aimed for the independent and structured assessment of patient indicators, we noticed that patients with complications seemed to report a poor or impaired status more frequently. Despite regular encouragement from researchers to fill in the diary, the responsiveness to it was limited. Interestingly, patients with events seemed less responsive to the diary compared with the control group, which might be a sign of deterioration by itself. Further investigation of patient-reported indicators for use in monitoring routines, perhaps using patient-tailored questionnaires or 2-way communication systems that promote patient response, is recommended.

Strengths and Limitations

This exploratory study included 2 surgical populations with a relatively high risk of developing postoperative complications and, therefore, may benefit from improved monitoring. The

XSL•FO
RenderX

heterogeneity of our study population and the variety of included events allowed the exploration of the predictive value of continuous EWSs for different algorithm settings. However, the total number of included patients and events eligible for analysis were limited, and it should be noted that clinical deterioration may present differently across other settings. Therefore, validation of this study's results in larger cohorts and other specific or general populations is needed once remote monitoring systems are mature and implemented more widely. Similarly, verification of the performance of high-rate EWSs for monitoring systems using other sensors or different parameter sets is desired. Finally, prospective studies assessing the clinical effects of using high-rate EWSs are needed.

The criteria used for the simulation of the sensor EWSs were based on clinical standards to allow comparison with current practice; therefore, the optimization of alarm criteria was limited to exploring different cutoff thresholds and sample frequencies. Furthermore, the EWSs were obtained using average vital sign values, whereas the use of other summary measures can result in different EWSs [44]. Because the results of this study revealed that sensor EWS–based alarms could lead to overdetection and high false alarm rates, further investigation of the best-suited preprocessing methods as well as alternative alarm criteria or methods is highly recommended; for example, it might be beneficial to personalize the allocation of EWS points by adapting normal ranges based on previous recordings of individuals or patient groups [24] or to correct for the influence of physical activity using embedded motion sensors. Moreover, because patterns of deterioration depend on the underlying cause [2] and patient characteristics, it could be useful to implement algorithms tailored to detect specific complications or patient groups [45]. Finally, trend-based metrics, machine learning techniques, or dynamic models may provide better performance for identifying abnormal patterns in continuous vital sign data compared with traditional EWSs, and event prediction may be improved further by integrating additional patient information extracted from the electronic patient record [16,28,45-51].

Although many studies have focused on the prediction of life-threatening events such as cardiac arrest, this study focused on all minor and major complications that required additional care in a ward setting. Patients with Clavien-Dindo class I complications that, by definition, did not require pharmacological or surgical treatment were excluded from the analysis. The sensitivity analysis indicated that the inclusion of these patients in the control group did not affect the results, indicating that the predictive value of vital signs or EWSs is limited for these events. Following the concept that early warning systems should alert only when action is required, this study used the onset of therapeutic actions as a surrogate marker of complication onset. However, the timing of therapeutic onset depends highly on the diagnostic definitions and is prone to delays in diagnostic or therapeutic activities. Furthermore, delays in clinical reporting hamper accurate retrospective determination of therapy onset. Therefore, prospective clinical trials are needed to investigate the effects of using a high-rate sensor-based EWSs as a diagnostic tool or alarm system on time to treatment and the corresponding patient outcomes.

Conclusions

Hourly EWSs obtained using wireless vital sign sensors resulted in higher sensitivity for predicting postoperative complications compared with the currently used 8-hour, nurse-based EWSs. Therefore, sensor EWSs may contribute to the improved detection of clinical deterioration in patients on surgical wards. In addition, using the EWS as an active alarm trigger results in fewer alarms compared with single-parameter alarms in sensor measurements. However, false alarm rates can increase rapidly when calculating EWSs in short time spans, risking alarm fatigue and overdetection. To prevent alarm overload in a ward setting, we suggest restricting EWS-based alarm generation to windows of at least 1 hour or longer. In addition, investigation of alternative care escalation criteria for wireless monitoring is warranted because automatic vital sign recordings differ from nurse measurements and are currently only available for a selected set of vital parameters. Finally, it is recommended to embed a structured collection of subjective nurse and patient information into monitoring routines, as this might provide complementary signs of deterioration. Future clinical studies are needed to evaluate the clinical effects of using the EWSs and the corresponding alarm system for continuous monitoring of patients on wards.

Acknowledgments

The authors thank all the patients who participated in the observational study and are grateful to the nurses and hospital staff who supported the study and data collection.

This study was supported by the Pioneers in Health Care Innovation fund (PIHC 2017).

Data Availability

The deidentified vital sign measurements that were analyzed during this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1

XSL•FO
RenderX

Nurse worry indicators.
[PDF File (Adobe PDF File), 125 KB - periop_v6i1e44483_app1.pdf]

Multimedia Appendix 2
Bland-Altman analysis.
[PDF File (Adobe PDF File), 223 KB - periop_v6i1e44483_app2.pdf]

References

1. Tevis SE, Kennedy GD. Postoperative complications and implications on patient-centered outcomes. J Surg Res 2013 May 01;181(1):106-113 [FREE Full text] [doi: 10.1016/j.jss.2013.01.032] [Medline: 23465392]
2. Thompson JS, Baxter BT, Allison JG, Johnson FE, Lee KK, Park WY. Temporal patterns of postoperative complications. Arch Surg 2003 Jun 01;138(6):596-603. [doi: 10.1001/archsurg.138.6.596] [Medline: 12799329]
3. Hyder JA, Wakeam E, Arora V, Hevelone ND, Lipsitz SR, Nguyen LL. Investigating the "rule of W," a mnemonic for teaching on postoperative complications. J Surg Educ 2015 May;72(3):430-437. [doi: 10.1016/j.jsurg.2014.11.004] [Medline: 25523129]
4. Amin T, Mobbs RJ, Mostafa N, Sy LW, Choy WJ. Wearable devices for patient monitoring in the early postoperative period: a literature review. Mhealth 2021 Jul 20;7:50 [FREE Full text] [doi: 10.21037/mhealth-20-131] [Medline: 34345627]
5. García-Del-Valle S, Arnal-Velasco D, Molina-Mendoza R, Gómez-Arnau JI. Update on early warning scores. Best Pract Res Clin Anaesthesiol 2021 May;35(1):105-113. [doi: 10.1016/j.bpa.2020.12.013] [Medline: 33742570]
6. Skervin A, Levy B. Management of common surgical complications. Surgery (Oxford) 2020 Mar;38(3):128-132 [FREE Full text] [doi: 10.1016/j.mpsur.2020.01.008]
7. Mok WQ, Wang W, Liaw SY. Vital signs monitoring to detect patient deterioration: an integrative literature review. Int J Nurs Pract 2015 May 26;21 Suppl 2:91-98 [FREE Full text] [doi: 10.1111/ijn.12329] [Medline: 26125576]
8. Odell M, Victor C, Oliver D. Nurses' role in detecting deterioration in ward patients: systematic literature review. J Adv Nurs 2009 Oct;65(10):1992-2006. [doi: 10.1111/j.1365-2648.2009.05109.x] [Medline: 20568317]
9. DeVita MA, Smith GB, Adam SK, Adams-Pizarro I, Buist M, Bellomo R, et al. "Identifying the hospitalised patient in crisis"--a consensus conference on the afferent limb of rapid response systems. Resuscitation 2010 Apr;81(4):375-382. [doi: 10.1016/j.resuscitation.2009.12.008] [Medline: 20149516]
10. Kellett J, Sebat F. Make vital signs great again - a call for action. Eur J Intern Med 2017 Nov;45:13-19. [doi: 10.1016/j.ejim.2017.09.018] [Medline: 28941841]
11. Eddahchouri Y, Koeneman M, Plokker M, Brouwer E, van de Belt TH, van Goor H, et al. Low compliance to a vital sign safety protocol on general hospital wards: a retrospective cohort study. Int J Nurs Stud 2021 Mar;115:103849 [FREE Full text] [doi: 10.1016/j.ijnurstu.2020.103849] [Medline: 33517078]
12. McGillion MH, Duceppe E, Allan K, Marcucci M, Yang S, Johnson AP, PROTECT Network Investigators. Postoperative remote automated monitoring: need for and state of the science. Can J Cardiol 2018 Jul;34(7):850-862 [FREE Full text] [doi: 10.1016/j.cjca.2018.04.021] [Medline: 29960614]
13. Leenen JPL, Leerentveld C, van Dijk JD, van Westreenen HL, Schoonhoven L, Patijn GA. Current evidence for continuous vital signs monitoring by wearable wireless devices in hospitalized adults: systematic review. J Med Internet Res 2020 Jun 17;22(6):e18636 [FREE Full text] [doi: 10.2196/18636] [Medline: 32469323]
14. Khanna AK, Hoppe P, Saugel B. Automated continuous noninvasive ward monitoring: future directions and challenges. Crit Care 2019 May 30;23(1):194 [FREE Full text] [doi: 10.1186/s13054-019-2485-7] [Medline: 31146792]
15. van Goor HM, van Loon K, Breteler MJ, Kalkman CJ, Kaasjager KA. Circadian patterns of heart rate, respiratory rate and skin temperature in hospitalized COVID-19 patients. PLoS One 2022 Jul 7;17(7):e0268065 [FREE Full text] [doi: 10.1371/journal.pone.0268065] [Medline: 35797369]
16. van Goor HM, Breteler MJ, Schoonhoven L, Kalkman CJ, van Loon K, Kaasjager KA. Interpretation of continuously measured vital signs data of COVID-19 patients by nurses and physicians at the general ward: a mixed methods study. PLoS One 2023 May 25;18(5):e0286080 [FREE Full text] [doi: 10.1371/journal.pone.0286080] [Medline: 37228047]
17. Leenen JP, Dijkman EM, van Dijk JD, van Westreenen HL, Kalkman C, Schoonhoven L, et al. Feasibility of continuous monitoring of vital signs in surgical patients on a general ward: an observational cohort study. BMJ Open 2021 Feb 17;11(2):e042735 [FREE Full text] [doi: 10.1136/bmjopen-2020-042735] [Medline: 33597138]
18. Subbe CP, Kruger M, Rutherford P, Gemmel L. Validation of a modified early warning Score in medical admissions. QJM 2001 Oct;94(10):521-526. [doi: 10.1093/qjmed/94.10.521] [Medline: 11588210]
19. Douw G, Huisman-de Waal G, van Zanten AR, van der Hoeven JG, Schoonhoven L. Nurses' 'worry' as predictor of deteriorating surgical ward patients: a prospective cohort study of the Dutch-Early-Nurse-Worry-Indicator-Score. Int J Nurs Stud 2016 Jul;59:134-140 [FREE Full text] [doi: 10.1016/j.ijnurstu.2016.04.006] [Medline: 27222458]
20. Dindo D, Demartines N, Clavien PA. Classification of surgical complications: a new proposal with evaluation in a cohort of 6336 patients and results of a survey. Ann Surg 2004 Aug;240(2):205-213. [doi: 10.1097/01.sla.0000133083.54934.ae] [Medline: 15273542]

XSL•FO
RenderX

21. National Early Warning Score (NEWS) 2: Standardising the Assessment of Acute-Illness Severity in the NHS. London, UK: Royal College of Physicians; 2017.
22. Bland JM, Altman DG. Agreement between methods of measurement with multiple observations per individual. J Biopharm Stat 2007 Jul 05;17(4):571-582 [FREE Full text] [doi: 10.1080/10543400701329422] [Medline: 17613642]
23. Weenk M, Koeneman M, van de Belt TH, Engelen LJ, van Goor H, Bredie SJ. Wireless and continuous monitoring of vital signs in patients at the general ward. Resuscitation 2019 Mar;136:47-53. [doi: 10.1016/j.resuscitation.2019.01.017] [Medline: 30685546]
24. van Rossum MC, Vlaskamp LB, Posthuma LM, Visscher MJ, Breteler MJ, Hermens HJ, et al. Adaptive threshold-based alarm strategies for continuous vital signs monitoring. J Clin Monit Comput 2022 Apr 11;36(2):407-417 [FREE Full text] [doi: 10.1007/s10877-021-00666-4] [Medline: 33575922]
25. Kause J, Smith G, Prytherch D, Parr M, Flabouris A, Hillman K, Intensive Care Society (UK), Australian and New Zealand Intensive Care Society Clinical Trials Group. A comparison of antecedents to cardiac arrests, deaths and emergency intensive care admissions in Australia and New Zealand, and the United Kingdom--the ACADEMIA study. Resuscitation 2004 Sep;62(3):275-282. [doi: 10.1016/j.resuscitation.2004.05.016] [Medline: 15325446]
26. Downey CL, Tahir W, Randell R, Brown JM, Jayne DG. Strengths and limitations of early warning scores: a systematic review and narrative synthesis. Int J Nurs Stud 2017 Nov;76:106-119 [FREE Full text] [doi: 10.1016/j.ijnurstu.2017.09.003] [Medline: 28950188]
27. Duus CL, Aasvang EK, Olsen RM, Sørensen HB, Jørgensen LN, Achiam MP, et al. Continuous vital sign monitoring after major abdominal surgery-quantification of micro events. Acta Anaesthesiol Scand 2018 Oct;62(9):1200-1208. [doi: 10.1111/aas.13173] [Medline: 29963706]
28. Churpek MM, Yuen TC, Winslow C, Meltzer DO, Kattan MW, Edelson DP. Multicenter comparison of machine learning methods and conventional regression for predicting clinical deterioration on the wards. Crit Care Med 2016 Feb;44(2):368-374 [FREE Full text] [doi: 10.1097/CCM.0000000000001571] [Medline: 26771782]
29. Haahr-Raunkjaer C, Mølgaard J, Elvekjaer M, Rasmussen SM, Achiam MP, Jorgensen LN, et al. Continuous monitoring of vital sign abnormalities; association to clinical complications in 500 postoperative patients. Acta Anaesthesiol Scand 2022 May;66(5):552-562 [FREE Full text] [doi: 10.1111/aas.14048] [Medline: 35170026]
30. Rothman MJ, Rothman SI, Beals JIV. Development and validation of a continuous measure of patient condition using the Electronic Medical Record. J Biomed Inform 2013 Oct;46(5):837-848 [FREE Full text] [doi: 10.1016/j.jbi.2013.06.011] [Medline: 23831554]
31. Itelman E, Shlomai G, Leibowitz A, Weinstein S, Yakir M, Tamir I, et al. Assessing the usability of a novel wearable remote patient monitoring device for the early detection of in-hospital patient deterioration: observational study. JMIR Form Res 2022 Jun 09;6(6):e36066 [FREE Full text] [doi: 10.2196/36066] [Medline: 35679119]
32. Subbe CP, Duller B, Bellomo R. Effect of an automated notification system for deteriorating ward patients on clinical outcomes. Crit Care 2017 Mar 14;21(1):52 [FREE Full text] [doi: 10.1186/s13054-017-1635-z] [Medline: 28288655]
33. Downey CL, Chapman S, Randell R, Brown JM, Jayne DG. The impact of continuous versus intermittent vital signs monitoring in hospitals: a systematic review and narrative synthesis. Int J Nurs Stud 2018 Aug;84:19-27 [FREE Full text] [doi: 10.1016/j.ijnurstu.2018.04.013] [Medline: 29729558]
34. van Rooijen CR, de Ruijter W, van Dam B. Evaluation of the threshold value for the Early Warning Score on general wards. Neth J Med 2013 Jan;71(1):38-43 [FREE Full text] [Medline: 23412825]
35. Green M, Lander H, Snyder A, Hudson P, Churpek M, Edelson D. Comparison of the between the flags calling criteria to the MEWS, NEWS and the electronic Cardiac Arrest Risk Triage (eCART) score for the identification of deteriorating ward patients. Resuscitation 2018 Feb;123:86-91 [FREE Full text] [doi: 10.1016/j.resuscitation.2017.10.028] [Medline: 29169912]
36. van Goor HM, Vernooij LM, Breteler MJ, Kalkman CJ, Kaasjager KA, van Loon K. Association of continuously measured vital signs with respiratory insufficiency in hospitalized COVID-19 patients: retrospective cohort study. Interact J Med Res 2022 Nov 23;11(2):e40289 [FREE Full text] [doi: 10.2196/40289] [Medline: 36256803]
37. van Rossum M, Leenen J, Kingma F, Breteler M, van Hillegersberg R, Ruurda J, et al. Expectations of continuous vital signs monitoring for recognizing complications after esophagectomy: interview study among nurses and surgeons. JMIR Perioper Med 2021 Feb 12;4(1):e22387 [FREE Full text] [doi: 10.2196/22387] [Medline: 33576743]
38. Breteler MJ, KleinJan EJ, Dohmen DA, Leenen LP, van Hillegersberg R, Ruurda JP, et al. Vital signs monitoring with wearable sensors in high-risk surgical patients: a clinical validation study. Anesthesiology 2020 Mar;132(3):424-439 [FREE Full text] [doi: 10.1097/ALN.0000000000003029] [Medline: 31743149]
39. Weenk M, van Goor H, Frietman B, Engelen LJ, van Laarhoven CJ, Smit J, et al. Continuous monitoring of vital signs using wearable devices on the general ward: pilot study. JMIR Mhealth Uhealth 2017 Jul 05;5(7):e91 [FREE Full text] [doi: 10.2196/mhealth.7208] [Medline: 28679490]
40. Haveman ME, van Rossum MC, Vaseur RM, van der Riet C, Schuurmann RC, Hermens HJ, et al. Continuous monitoring of vital signs with wearable sensors during daily life activities: validation study. JMIR Form Res 2022 Jan 07;6(1):e30863 [FREE Full text] [doi: 10.2196/30863] [Medline: 34994703]

XSL•FO
RenderX

41. Stellpflug C, Pierson L, Roloff D, Mosman E, Gross T, Marsh S, et al. Continuous physiological monitoring improves patient outcomes. Am J Nurs 2021 Apr 01;121(4):40-46. [doi: 10.1097/01.NAJ.0000742504.44428.c9] [Medline: 33755624]
42. Romero-Brufau S, Gaines K, Nicolas CT, Johnson MG, Hickman J, Huddleston JM. The fifth vital sign? Nurse worry predicts inpatient deterioration within 24 hours. JAMIA Open 2019 Dec;2(4):465-470 [FREE Full text] [doi: 10.1093/jamiaopen/ooz033] [Medline: 32025643]
43. Rainey H, Ehrich K, Mackintosh N, Sandall J. The role of patients and their relatives in 'speaking up' about their own safety - a qualitative study of acute illness. Health Expect 2015 Jun;18(3):392-405 [FREE Full text] [doi: 10.1111/hex.12044] [Medline: 23332029]
44. Youssef Ali Amer A, Wouters F, Vranken J, de Korte-de Boer D, Smit-Fun V, Duflot P, et al. Vital signs prediction and early warning score calculation based on continuous monitoring of hospitalised patients using wearable technology. Sensors (Basel) 2020 Nov 18;20(22):6593 [FREE Full text] [doi: 10.3390/s20226593] [Medline: 33218084]
45. Petit C, Bezemer R, Atallah L. A review of recent advances in data analytics for post-operative patient deterioration detection. J Clin Monit Comput 2018 Jun;32(3):391-402. [doi: 10.1007/s10877-017-0054-7] [Medline: 28828569]
46. Hravnak M, Pellathy T, Chen L, Dubrawski A, Wertz A, Clermont G, et al. A call to alarms: current state and future directions in the battle against alarm fatigue. J Electrocardiol 2018 Nov;51(6S):S44-S48 [FREE Full text] [doi: 10.1016/j.jelectrocard.2018.07.024] [Medline: 30077422]
47. Joshi M, Ashrafian H, Aufegger L, Khan S, Arora S, Cooke G, et al. Wearable sensors to improve detection of patient deterioration. Expert Rev Med Devices 2019 Mar 06;16(2):145-154 [FREE Full text] [doi: 10.1080/17434440.2019.1563480] [Medline: 30580650]
48. Mann KD, Good NM, Fatehi F, Khanna S, Campbell V, Conway R, et al. Predicting patient deterioration: a review of tools in the digital hospital setting. J Med Internet Res 2021 Sep 30;23(9):e28209 [FREE Full text] [doi: 10.2196/28209] [Medline: 34591017]
49. Muralitharan S, Nelson W, Di S, McGillion M, Devereaux PJ, Barr NG, et al. Machine learning-based early warning systems for clinical deterioration: systematic scoping review. J Med Internet Res 2021 Feb 04;23(2):e25187 [FREE Full text] [doi: 10.2196/25187] [Medline: 33538696]
50. Zhu Y, Chiu YD, Villar SS, Brand JW, Patteril MV, Morrice DJ, et al. Dynamic individual vital sign trajectory early warning score (DyniEWS) versus snapshot national early warning score (NEWS) for predicting postoperative deterioration. Resuscitation 2020 Dec;157:176-184 [FREE Full text] [doi: 10.1016/j.resuscitation.2020.10.037] [Medline: 33181231]
51. Gu Y, Rasmussen S, Mølgaard J, Haahr-Raunkjær C, Meyhoff CS, Aasvang EK, et al. Prediction of severe adverse event from vital signs for post-operative patients. Annu Int Conf IEEE Eng Med Biol Soc 2021 Nov;2021:971-974. [doi: 10.1109/EMBC46164.2021.9630918] [Medline: 34891450]

Abbreviations

AT: axillary temperature
bpm: beats per minute
brpm: breaths per minute
EWS: early warning score
FDR: false discovery rate
FP: false-positive
GI: gastrointestinal
HR: heart rate
MEWS: Modified Early Warning Score
nurse EWS: early warning score based on manual vital signs measurements
RR: respiratory rate
sensor EWS: early warning score based on sensor vital signs measurements
SpO_2: blood oxygen saturation
TAR: total alarm rate
TP: true-positive

XSL•FO
RenderX

Edited by T Leung, T de Azevedo Cardoso; submitted 24.11.22; peer-reviewed by M Binandeh, MCS Wong; comments to author 08.03.23; revised version received 16.06.23; accepted 07.07.23; published 30.08.23.

Please cite as:
van Rossum MC, Bekhuis REM, Wang Y, Hegeman JH, Folbert EC, Vollenbroek-Hutten MMR, Kalkman CJ, Kouwenhoven EA, Hermens HJ
Early Warning Scores to Support Continuous Wireless Vital Sign Monitoring for Complication Prediction in Patients on Surgical Wards: Retrospective Observational Study
JMIR Perioper Med 2023;6:e44483
URL: https://periop.jmir.org/2023/1/e44483
doi:10.2196/44483
PMID:37647104

©Mathilde C van Rossum, Robin E M Bekhuis, Ying Wang, Johannes H Hegeman, Ellis C Folbert, Miriam M R Vollenbroek-Hutten, Cornelis J Kalkman, Ewout A Kouwenhoven, Hermie J Hermens. Originally published in JMIR Perioperative Medicine (http://periop.jmir.org), 30.08.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Perioperative Medicine, is properly cited. The complete bibliographic information, a link to the original publication on http://periop.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Efficacy of Electronic Reminders in Increasing the Enhanced Recovery After Surgery Protocol Use During Major Breast Surgery: Prospective Cohort Study

Sumeet Gopwani[1*], MD; Ehab Bahrun[2*], BS; Tanvee Singh[2*], BS; Daniel Popovsky[2*], MS; Joseph Cramer[3*], MD; Xue Geng[4*], MS

[1]Department of Anesthesiology, MedStar Georgetown University Hospital, Washington, DC, United States

[2]Georgetown University School of Medicine, Washington, DC, United States

[3]Department of Anesthesiology, Perioperative and Pain Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[4]Department of Biostatistics, Bioinformatics & Biomathematics, Georgetown University, Washington, DC, United States

[*]all authors contributed equally

Corresponding Author:
Sumeet Gopwani, MD
Department of Anesthesiology
MedStar Georgetown University Hospital
3800 Reservoir Road NW
Washington, DC, 20007
United States
Phone: 1 7344764041
Email: sumeet.gopwani@medstar.net

Abstract

Background: Enhanced recovery after surgery (ERAS) protocols are patient-centered, evidence-based guidelines for peri-, intra-, and postoperative management of surgical candidates that aim to decrease operative complications and facilitate recovery after surgery. Anesthesia providers can use these protocols to guide decision-making and standardize aspects of their anesthetic plan in the operating room.

Objective: Research across multiple disciplines has demonstrated that clinical decision support systems have the potential to improve protocol adherence by reminding providers about departmental policies and protocols via notifications. There remains a gap in the literature about whether clinical decision support systems can improve patient outcomes by improving anesthesia providers' adherence to protocols. Our hypothesis is that the implementation of an electronic notification system to anesthesia providers the day prior to scheduled breast surgeries will increase the use of the already existing but underused ERAS protocols.

Methods: This was a single-center prospective cohort study conducted between October 2017 and August 2018 at an urban academic medical center. After obtaining approval from the institutional review board, anesthesia providers assigned to major breast surgery cases were identified. Patient data were collected pre- and postimplementation of an electronic notification system that sent the anesthesia providers an email reminder of the ERAS breast protocol the night before scheduled surgeries. Each patient's record was then reviewed to assess the frequency of adherence to the various ERAS protocol elements.

Results: Implementation of an electronic notification significantly improved overall protocol adherence and several preoperative markers of ERAS protocol adherence. Protocol adherence increased from 16% (n=14) to 44% (n=44; $P<.001$), preoperative administration of oral gabapentin (600 mg) increased from 13% (n=11) to 43% (n=43; $P<.001$), and oral celebrex (400 mg) use increased from 16% (n=14) to 35% (n=35; $P=.006$). There were no statistically significant differences in the use of scopolamine transdermal patch ($P=.05$), ketamine ($P=.35$), and oral acetaminophen ($P=.31$) between the groups. Secondary outcomes such as intraoperative and postoperative morphine equivalent administered, postanesthesia care unit length of stay, postoperative pain scores, and incidence of postoperative nausea and vomiting did not show statistical significance.

Conclusions: This study examines whether sending automated notifications to anesthesia providers increases the use of ERAS protocols in a single academic medical center. Our analysis exhibited statistically significant increases in overall protocol adherence but failed to show significant differences in secondary outcome measures. Despite the lack of a statistically significant difference in secondary postoperative outcomes, our analysis contributes to the limited literature on the relationship between using push

XSL•FO
RenderX

notifications and clinical decision support in guiding perioperative decision-making. A variety of techniques can be implemented, including technological solutions such as automated notifications to providers, to improve awareness and adherence to ERAS protocols.

(JMIR Perioper Med 2023;6:e44139) doi:10.2196/44139

KEYWORDS

ERAS protocol; electronic notification system; clinical decision support system; postoperative outcomes; breast surgery; surgery; surgical; postoperative; decision support; notification; recovery; anesthesia; cohort study; patient outcome; enhanced recovery; patient education; surgical stress

Introduction

Standardizing practices in a variety of industries has proven to be effective in the productivity and quality of work. With respect to perioperative care, a major development of standardized evidence-based practice is implementing clinical guidelines and protocols, such as enhanced recovery after surgery (ERAS) protocols [1]. ERAS protocols are multidisciplinary evidence-based guidelines for the care of surgical patients. The aim is to minimize the stress of surgery and support patients to recover as soon as possible. ERAS protocols have increasingly been shown to improve postoperative outcomes for patients in a variety of surgical procedures [2,3]. In particular, it has been found to reduce hospital stay lengths and improve patient outcomes in a cost-effective manner [2]. Protocol use in the health care setting is increasingly seen as a driver of quality and safety via a standardization of practice. However, adherence to ERAS protocols by providers has been a challenge. One reason could be that ERAS protocols are considered complex and resource-demanding [2]. Introducing protocols usually requires a major shift in clinical practice and many health care providers have difficulties making these changes [2].

Fortunately, studies have demonstrated that clinical decision support systems (CDSS) for anesthesia have the potential to improve protocol adherence by reminding providers about departmental policies and protocols [1-5]. A CDSS system consists of three modules: (1) data acquisition through the electronic health record or anesthesia information management system, (2) processing of data through rules modulation, and (3) notification to the health care provider [6]. Several examples of CDSS systems can be noted within anesthesiology and perioperative management, such as devices that alert clinicians to abnormalities in blood pressure readings, or systems that warn of excessive anesthetic being administered [6]. It is clear that CDSS systems offer many applications in knowledge management and passive provider education [6]. CDSS have seen a rapid evolution since their first use in the 1980s. These systems are now commonly administered through electronic medical records or other computerized clinical workflows [6]. Several industries across the services and technology sector have implemented the use of push notifications as a primary method to improve communications to targeted users. Notification systems have also been implemented within the health care setting as a means of CDSS [7]. Here, we examine the use of these push notifications to increase ERAS protocol adherence by providers caring for patients undergoing major breast surgery. Our hypothesis is that the implementation of an electronic notification system to anesthesia providers the day prior to scheduled breast surgeries will increase the use of the ERAS protocols. We hypothesize that this notification system will draw providers' attention to the protocol enhancing use and improving patient outcomes.

Methods

Recruitment

This was a single-center prospective cohort study conducted between October 2017 and August 2018 at MedStar Georgetown University Hospital, a tertiary academic medical center in Washington DC. Participants met the inclusion criteria if they were attending anesthesiologists, anesthesiology residents, certified registered nurse anesthetists, and student registered nurse anesthetists in the department assigned to major breast surgery cases. Attending anesthesiologists involved in writing the ERAS protocol, and anesthesia providers in this study were excluded from participation. For the purposes of this study, mastectomies and breast reconstruction such as deep inferior epigastric perforator flaps were the only 2 surgical procedures that were considered major breast surgeries. Other procedures performed by breast surgeons were not included in this analysis.

The current focus of ERAS at our institution is preoperative and intraoperative interventions such as premedication and anesthetic technique, that is, avoiding narcotics. For the purpose of this study, an increase in guideline adherence was defined as a 20% increase in the use of scopolamine, gabapentin, and celecoxib. The specific protocol can be seen in Multimedia Appendix 1. Previous studies of ERAS protocols have used different rates to exhibit adherence. A 20% increase in protocol use was set due to previous research using this standard to define adherence [8]. The preintervention database included 100 patients who underwent major breast surgery from October 2017 to January 2018. Similarly, the postintervention database included 100 patients who underwent major breast surgery from February to August 2018. Based on this study's power analysis, 83 patients in each category were needed to detect the predefined outcome with statistical significance. The sample size needed to achieve adequate power, in this case, set at 80%, was done using a web-based clinical calculator called ClinCalc. This calculator uses study group design, primary end point, and statistical parameters such as anticipated incidence and type 1 and 2 error rate in order to calculate the size needed to achieve a set power value.

Using an institutional electronic medical record, a database was created, which identified patients who were undergoing major

XSL•FO
RenderX

breast surgery. Patient data were anonymous and deidentified. For the postintervention group, practitioners assigned to operating rooms with these surgeries were also identified the day prior to surgery, and the following notification was sent to the respective anesthesia provider via email:

> *We have identified that you will be in breast surgery tomorrow, where the attached Early Recovery after Surgery (ERAS) protocol may or may not be applicable. Please consider whether this protocol is appropriate for each of your patients individually. The attending anesthesiologist will make the final determination of appropriate care for each patient.*

The purpose of electronic notifications was to remind providers that ERAS protocols are in place rather than have them use traditional human and system factors. Electronic notifications were emailed to providers in an automated fashion at 7 PM every night prior to surgery using a CDSS built by the study team. The CDSS included data acquisition, data processing, and provider notification modules. Anesthesiology staff assignments and operating room schedules are uploaded as PDF documents onto a WordPress-based departmental website. Data acquisition was done with a custom-designed AppleScript to extract the schedules from the website each evening and process them through optical character recognition software (Adobe Acrobat Pro DC). In the Apple Xcode (Apple, Cupertino) development environment, using the C++ programming language, an algorithm was written to process the schedules, parsing the text with a delimiter function. This allowed the identification of providers assigned to major breast surgery. A notification script (AppleScript) was implemented to push emails to providers with instructions as well as the ERAS protocol for breast surgery. Anesthesia providers used the ERAS protocol contained within the email and did not use any premade order sets.

Data were collected from the patient's electronic medical chart in both the pre- and postintervention groups including the date and type of surgery as well as the type of anesthesia provided. For the purposes of this study, the focus was on the medications that were administered preoperatively, intraoperatively, and postoperatively. This included the use of gabapentin, celecoxib, acetaminophen, ketamine, scopolamine, fentanyl, and ondansetron.

The primary outcome of this study was to measure the adherence to the ERAS protocol, which is determined by preoperative and intraoperative use of medications. Preoperative medications used were as follows: oral gabapentin 600 mg, oral celebrex 400 mg, oral acetaminophen 1000 mg for AM admissions, and transdermal scopolamine patch. Intraoperative medications used were as follows: midazolam PRN, propofol gtt, ketamine 20 mg IV ± gtt at 0.2 mg/kg/h, decadron 8 mg IV at induction, ondansetron 4 mg IV prior to extubation, and acetaminophen 1000 mg IV if not given preoperatively. Secondary outcomes included intraoperative and postoperative narcotics administered, postanesthesia care unit (PACU) length of stay (LOS), first pain score, highest pain scores postoperatively, and incidence of postoperative nausea and vomiting (PONV). Narcotics used in this study included: fentanyl, morphine, dilaudid, and oxycodone, which were considered to be morphine equivalents. In the ERAS protocol sent to providers, there was a distinct pathway that detailed the following stages: preoperative clinic, preoperative holding, intraoperative, and postoperative stages. Each patient care stage was clearly labeled with the appropriate pharmacologic interventions required, specifically, what dosage of the drugs previously mentioned was appropriate and whether the drug was administered intravenously or orally.

Ethical Considerations

Approval was obtained from the MedStar Health Research Institute institutional review board (STUDY #2017-0725), and informed consent was obtained from all patients. Secondary analysis of data was allowed per MedStar Health institutional review board protocol after obtaining initial primary consent. Retrospective data were analyzed in accordance with MedStar Health's nonhuman subjects research policies.

Statistical Analysis

All statistical analyses were performed using the statistical software RStudio (version 1.4.1106; Posit). For this study, we accepted a P value less than .05 for statistical significance. The initial data included 200 participants, with 100 in the prenotification group and 100 in the postnotification group. Data characteristics were summarized by frequency and percentage for categorical variables and mean and SD or median and IQR for continuous variables based on the normality of the data. Shapiro-Wilk test was used to check the normality of the continuous variables. The following variables were used in analyses between groups: patient age, weight, American Society of Anesthesiology (ASA) score, PACU LOS, first pain score, highest pain score, intraoperative morphine equivalent, and postoperative morphine equivalent. Student 2-sided t test or Wilcoxon rank-sum test was used to check whether there was a significant association between continuous variables and categorical variables based on the normality of the continuous variables. Chi-square test was used to determine whether there was a significant association between 2 categorical variables, and Fisher exact test was used instead of chi-square test if a cell count was less than 5.

Results

Participants' Demographic Information

In total, 12 patients in the preintervention group who had minor breast surgery were excluded from all analyses. Additionally, 3 patients from the preintervention group and 1 patient from the postintervention group were also excluded from all analyses due to missing data. Demographic variables of the patients undergoing breast surgery are shown in Table 1. The variable patient age was normally distributed, whereas the following 7 continuous variables were not distributed normally: weight, ASA score, PACU LOS, first pain score, highest pain score, intraoperative morphine equivalent, and postoperative morphine equivalent. A total of 176 patient cases were included in this study; 85 cases where anesthesia providers were in the pre-electronic notification group and 99 cases where the anesthesia providers were in the postelectronic notification group. Of the 85 participants providing anesthetic care in the

XSL•FO
RenderX

pre-electronic notification group, the mean age of patients undergoing breast surgery was 50.81 (SD 12.65) years, the median weight was 70.8 (IQR 61.50-84.00) kg, and the median ASA score was 2.00 (IQR 2.00-3.00). Of the 99 participants providing anesthetic care in the postelectronic notification group, the mean age of patients undergoing breast surgery was 50.74 (SD 14.30) years, the median weight was 76 (IQR 61.35-87.60) kg, and the median ASA score was 2.00 (IQR 2.00-3.00). There were no statistically significant differences in mean age, median weight, and median ASA score (*P*>.05) between the groups.

Table 1. Baseline characteristics of patients included in the study (N=176).

Characteristic	Pre-electronic notification	Postelectronic notification	*P* value
Age (years), mean (SD)	50.81 (12.65)	50.74 (14.30)	.97
Weight (kg), median (IQR)	70.8 (61.50-84.00)	76.00 (61.35-87.60)	.25
ASA[a] score, median (IQR)	2.00 (2.00-3.00)	2.00 (2.00-3.00)	.82

[a]ASA: American Society of Anesthesiology.

Increased Adherence to the ERAS Protocol

Implementation of an electronic notification significantly improved overall protocol adherence, and several preoperative markers of ERAS protocol adherence are shown in Table 2. The overall protocol adherence, use of oral gabapentin (600 mg), and oral celecoxib (400 mg) showed a statistically significant increase in use (Table 2). With respect to overall protocol adherence, 17% (14/85) of patients followed protocol in the pre-electronic notification group compared to 44% (44/99) of patients in the postelectronic notification group (*P*<.001). Among those in the pre-electronic notification group, 11% (11/85) received gabapentin, whereas in the postelectronic notification group, 43% (43/99) received gabapentin (*P*<.001). Lastly, 14 patients received celecoxib in the prenotification group, whereas 35 received it in the postnotification group (*P*=.006). There were no statistically significant differences in the use of scopolamine transdermal patch (*P*=.05), intraoperative ketamine (*P*=.36), and oral acetaminophen (*P*=.31) between the groups.

Table 2. Impact on each element of protocol.

Intervention	Pre-electronic notification, n (%)	Postelectronic notification, n (%)	*P* value
Protocol adherence	14 (17)	44 (44)	<.001
Gabapentin	11 (13)	43 (43)	<.001
Scopolamine	25 (29)	44 (44)	.05
Ketamine	3 (4)	7 (7)	.35
Celecoxib	14 (17)	35 (35)	.006
Acetaminophen	7 (8)	14 (14)	.31

Secondary Outcomes

Next, we analyzed postoperative outcomes of patients undergoing breast surgery stratifying according to whether their anesthesia provider received an electronic notification for the ERAS protocol (Table 3). Secondary outcomes were intraoperative and postoperative morphine equivalent administered, PACU LOS, postoperative pain scores, and incidence of PONV. We did not show any statistically significant improvements in secondary patient outcomes.

Table 3. Secondary outcomes analysis.

Postoperative outcomes	Pre-electronic notification	Postelectronic notification	*P* value
Intraoperative morphine equivalent, median (IQR)	20.00 (10.00-30.00)	20.00 (10.00-30.00)	.55
Postoperative morphine equivalent, median (IQR)	6.11 (1.11-11.11)	6.11 (2.22-12.22)	.70
PACU[a] LOS[b], median (IQR)	118.00 (92.00-162.00)	154.00 (84.00-217.50)	.29
First pain score, median (IQR)	4.00 (2.00-6.00)	4.00 (2.00-6.00)	.39
Highest pain score, median (IQR)	5.00 (4.00-7.00)	5.00 (3.00-7.00)	.95
PONV[c], n (%)	27 (32)	39 (39.4)	.36

[a]PACU: postanesthesia care unit.

[b]LOS: length of stay.

[c]PONV: postoperative nausea and vomiting.

XSL•FO
RenderX

The median postoperative morphine equivalent was 6.11 (IQR 1.11-11.11) for the prenotification group (85 patients) and 6.11 (IQR 2.22-12.22) for the postnotification group (99 patients), and there was no statistically significant difference in postoperative morphine equivalent between prenotification and postnotification groups (*P*=.70). The median intraoperative morphine equivalent was 20.00 (IQR 10.00-30.00) for both the prenotification and the postnotification groups of patients, and there was no statistically significant difference in intraoperative morphine equivalent between prenotification and postnotification groups (*P*=.55). LOS in the PACU was not statistically significant between the 2 groups, with the pre-electronic notification group having a median of 118.00 (IQR 92.00-162.00) minutes and postelectronic notification group having a median of 154.00 (IQR 84.00-217.50) minutes (*P*=.29). The median of the first recorded pain score was 4.00 (IQR 2.00-6.00) for both the pre-electronic notification group and postelectronic notification group (*P*=.31). The median highest recorded pain score was 5.00 (IQR 4.00-7.00) for the pre-electronic notification group and 5.00 (IQR 3.00-7.00) for the postelectronic notification group (*P*=.95). Lastly, the incidence of PONV was 27 (32%) among the pre-electronic notification group and 39 (39%) among the postelectronic notification group (*P*=.36).

Discussion

Principal Findings

This study exhibited statistically significant primary outcome measures estimating adherence to ERAS protocols determined by pre- and intraoperative use of medications but did not exhibit statistically significant secondary outcome measures. Recent studies have demonstrated benefits with the implementation of ERAS protocols in a variety of surgeries including cardiothoracic, gastrointestinal, and gynecologic surgery. Benefits of ERAS protocols included shorter LOS, more rapid return of bowel function, decreased postoperative pain, and increased patient satisfaction [2,3,9]. With respect to breast surgery, ERAS protocols have been shown to reduce the length of hospital stay, opioid use, and PONV without increasing the rates of complication [4,10-12].

Previous studies have also demonstrated the importance of decision support systems (DSS) in the clinical setting. The purpose of DSS is to aid clinicians in centralizing the increasing amounts of data for each patient alongside the increasing volume of medical research [13]. There are several types of CDSS often categorized based on the following characteristics: system function, model for advice, human interaction, and underlying decision-making process [14]. Our method of CDSS closely resembled the model for advice. This model, however, can be classified into 2 subcategories: passive and active. Passive DSS require the user to perform an action to receive advice. Active DSS, the model used in our study, involve the generation of alerts to providers as a means of decision support [14]. The latter has been shown to be efficacious in the perioperative phase. For instance, Kooij et al [15] demonstrated how electronic CDSS increased guidelines adherence for prescribing PONV prophylaxis. Our study sought to further investigate the benefits of CDSS in implementing ERAS protocols in patients undergoing major breast surgery.

Our analysis showed an improvement in multiple preoperative markers of the ERAS protocols suggesting improved guideline adherence. For instance, there was an increase in the use of gabapentin and celecoxib. This statistically significant difference from our pre- and postintervention groups supports our primary hypothesis that an electronic notification system will impact the frequency of which providers incorporate the breast surgery ERAS protocols. Other studies have shown similar results. One study examined the implementation of electronic alerts for improving adherence to foot exam screenings in type 2 diabetic patients in primary care clinics [16]. The researchers demonstrated that the use of an electronic clinical reminder to providers increased adherence and subsequently resulted in clinically significant outcomes. The aid of ERAS protocols has also been investigated in the perioperative setting where electronic DSS increased guideline adherence for the prescription of PONV prophylaxis and in the intraoperative setting where real-time electronic reminders improved compliance to institutional glucose management [14,17].

Although there was a statistically significant increase in the use of gabapentin and celecoxib, our analysis showed no statistically significant difference in the use of transdermal scopolamine, ketamine, or acetaminophen in the pre- and postintervention groups. This can be due to several reasons. For instance, with ketamine, there is limited evidence of its use in breast surgery [18]. In a randomized controlled trial evaluating for this effect, patients undergoing mastectomies were randomized to receive ketamine (0.15 mg/kg IV) before surgery or during closure. The researchers concluded that ketamine at the end of surgery was more effective in reducing morphine consumption as patient-controlled analgesia was lower during the first 2 hours in patients given ketamine at closure [18]. With the abundance of evidence that ketamine provides preoperative analgesia for multiple surgical procedures, it is probable that ketamine may be beneficial for breast surgery. However, a dearth of evidence specific to breast surgery may cause anesthesiology providers to preclude it from their practice [18].

With respect to acetaminophen, the lack of a statistically significant difference is likely due to changes in the preferred route of administration over the past few years. During the time of this study, intravenous acetaminophen was perceived as the gold standard for multimodal pain management at our institution. Consequently, only a minority of our patients, less than 15%, in both the pre- and postintervention groups, received oral acetaminophen as recommended per the ERAS protocols. However, recent studies have shown oral acetaminophen given preoperatively was equivalent to intravenous acetaminophen in controlling pain in the immediate postoperative phase [19]. Intravenous acetaminophen was also not found to be superior to oral acetaminophen in reducing time to ambulation, length of PACU stay, or PONV [19]. Consequently, intravenous acetaminophen was not included in our adherence calculation and only oral dosing was included as a factor for adherence. Given the clinical practice changes at our institution over recent years, if our study was done in the present, it is likely that the vast majority of our patients would have received oral

XSL•FO
RenderX

acetaminophen, although the presence of a statistically significant difference between the pre- and postintervention groups would still be in question.

Some studies have reported that the adoption of clinical decision support tools could have a significant impact on the performance of providers [20]. However, it is erroneous to establish an equivalence between provider behaviors and patient outcomes. Research has shown it is imperative to assess the impact of clinical decision support on provider behavior and objective clinical end points [5]. Additionally, it is important to note DSS are only 1 component of increasing compliance and that increasing guideline adherence is multifactorial. This study shows that CDSS can be a meaningful component of a multifactorial protocol adherence strategy. Using traditional human and system factors to enforce clinical guidelines often results in suboptimal guideline adherence [21]. The goal of our electronic notification system was to augment the traditional systems in place that remind providers about existing protocols, in this case, an ERAS protocol for breast surgery. Our study also adds to the existing literature by analyzing secondary patient outcomes such as PACU LOS, morphine dosage, pain scores, and PONV stratifying according to whether the anesthesia provider received an electronic notification for the ERAS protocol. Interestingly, none of these variables showed statistical significance between the pre- and postnotification groups. This finding is consistent with the existing literature on perioperative clinical decision support. As explained by Nair et al [22], these are complex multifactorial perioperative outcomes, and CDSS may represent merely 1 piece of a multifaceted quality improvement strategy.

Limitations

The findings of this study should be interpreted in the context of the following limitations. Data are limited on variables such as the length of surgery, which could plausibly introduce bias. We also had to exclude 12 patients in the prenotification group that had minor breast surgery (ie, breast biopsy and lumpectomies). These 12 patients were also excluded from the secondary analysis, as well as excluding other 3 patients from the preintervention group and 1 patient from the postintervention group with missing data. Nonetheless, secondary postoperative patient outcomes were not found to be significant. Adherence to protocol was defined as a 20% increase in protocol use. Unfortunately, there is no defined standard to quantify adherence, and there is widespread variation of this metric across literature, but for the purpose of our study, we wanted to show an incremental increase in usage, knowing that gauging adherence is a multimodal approach. Furthermore, these findings are from data collected in a single large academic medical center and may not be representative of a larger cohort. Future studies, ideally multicenter, are needed to establish a more robust relationship between electronic notification systems and provider adherence.

Conclusions

This study exhibited statistically significant primary outcome measures measuring adherence to ERAS protocols determined by pre- and intraoperative use of medications but did not exhibit statistically significant secondary outcome measures including intraoperative and postoperative narcotic administered, PACU LOS, first pain score, highest pain scores postoperatively, and incidence of PONV. Although existing studies have identified the benefits of ERAS protocols, several barriers are present, which prevent the practice of these guidelines. Our study demonstrates that 1 technique to overcome this is by using automated notifications for anesthesia care providers. This paper contributes to the existing literature by examining how electronic notifications could increase adherence to ERAS protocol use and shows how one can build a relatively simple CDSS using our methodology to do so. In the future, we hope to see this model repeated by anesthesiologists across the country with various alterations to examine other areas in which patients may benefit from decision support systems.

Conflicts of Interest

None declared.

Multimedia Appendix 1
ERAS protocol used for breast surgery sent to surgeons and anesthesiologists.
[PDF File (Adobe PDF File), 149 KB - periop_v6i1e44139_app1.pdf]

References

1. Boudreaux P. Barriers and facilitators influencing compliance with enhanced recovery after surgery protocol: a qualitative study. LSU Master's Theses. 2020. URL: https://digitalcommons.lsu.edu/gradschool_theses/5085/ [accessed 2021-07-16]
2. Gustafsson UO, Hausel J, Thorell A, Ljungqvist O, Soop M, Nygren J, Enhanced Recovery After Surgery Study Group. Adherence to the enhanced recovery after surgery protocol and outcomes after colorectal cancer surgery. Arch Surg 2011;146(5):571-577 [FREE Full text] [doi: 10.1001/archsurg.2010.309] [Medline: 21242424]
3. Pędziwiatr M, Mavrikis J, Witowski J, Adamos A, Major P, Nowakowski M, et al. Current status of enhanced recovery after surgery (ERAS) protocol in gastrointestinal surgery. Med Oncol 2018;35(6):95 [FREE Full text] [doi: 10.1007/s12032-018-1153-0] [Medline: 29744679]
4. Ljungqvist O, Scott M, Fearon KC. Enhanced recovery after surgery: a review. JAMA Surg 2017;152(3):292-298. [doi: 10.1001/jamasurg.2016.4952] [Medline: 28097305]

XSL•FO
RenderX

5. Garg AX, Adhikari NKJ, McDonald H, Rosas-Arellano MP, Devereaux PJ, Beyene J, et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. JAMA 2005;293(10):1223-1238. [doi: 10.1001/jama.293.10.1223] [Medline: 15755945]
6. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digit Med 2020;3:17 [FREE Full text] [doi: 10.1038/s41746-020-0221-y] [Medline: 32047862]
7. Kheterpal S, Shanks A, Tremper KK. Impact of a novel multiparameter decision support system on intraoperative processes of care and postoperative outcomes. Anesthesiology 2018;128(2):272-282 [FREE Full text] [doi: 10.1097/ALN.0000000000002023] [Medline: 29337743]
8. Springer JE, Doumouras AG, Lethbridge S, Forbes S, Eskicioglu C. The predictors of enhanced recovery after surgery utilization and practice variations in elective colorectal surgery: a provincial survey. Can J Surg 2020;63(5):E460-E467 [FREE Full text] [doi: 10.1503/cjs.009419] [Medline: 33107814]
9. American College of Obstetricians and Gynecologists. ACOG committee opinion no. 750: perioperative pathways: enhanced recovery after surgery. Obstet Gynecol 2018;132(3):e120-e130 [FREE Full text] [doi: 10.1097/AOG.0000000000002818] [Medline: 30134426]
10. Chiu C, Aleshi P, Esserman LJ, Inglis-Arkell C, Yap E, Whitlock EL, et al. Improved analgesia and reduced post-operative nausea and vomiting after implementation of an enhanced recovery after surgery (ERAS) pathway for total mastectomy. BMC Anesthesiol 2018;18(1):41 [FREE Full text] [doi: 10.1186/s12871-018-0505-9] [Medline: 29661153]
11. Jogerst K, Thomas O, Kosiorek HE, Gray R, Cronin P, Casey W, et al. Same-day discharge after mastectomy: breast cancer surgery in the Era of ERAS. Ann Surg Oncol 2020;27(9):3436-3445. [doi: 10.1245/s10434-020-08386-w] [Medline: 32221736]
12. Offodile AC, Gu C, Boukovalas S, Coroneos CJ, Chatterjee A, Largo RD, et al. Enhanced recovery after surgery (ERAS) pathways in breast reconstruction: systematic review and meta-analysis of the literature. Breast Cancer Res Treat 2019;173(1):65-77. [doi: 10.1007/s10549-018-4991-8] [Medline: 30306426]
13. Pusic M, Ansermino D. Clinical decision support systems. B C Med J 2004;46(5):236-239 [FREE Full text]
14. Wasylewicz A, Scheepers-Hoeks AMJW. Clinical decision support systems. In: Kubben P, Dumontier M, Dekker A, editors. Fundamentals of Clinical Data Science. Cham, Switzerland: Springer; 2018.
15. Kooij FO, Klok T, Hollmann MW, Kal JE. Decision support increases guideline adherence for prescribing postoperative nausea and vomiting prophylaxis. Anesth Analg 2008;106(3):893-898 [FREE Full text] [doi: 10.1213/ane.0b013e31816194fb] [Medline: 18292437]
16. Denson R. Implementation of an electronic alert for improving adherence to diabetic foot exam screenings in type 2 diabetic patients in primary care clinics. Student Scholarly Projects, University of St Augustine for Health Sciences. 2020. URL: https://soar.usa.edu/cgi/viewcontent.cgi?article=1020&context=scholprojects [accessed 2021-05-08]
17. Nair BG, Grunzweig K, Peterson GN, Horibe M, Neradilek MB, Newman SF, et al. Intraoperative blood glucose management: impact of a real-time decision support system on adherence to institutional protocol. J Clin Monit Comput 2016;30(3):301-312. [doi: 10.1007/s10877-015-9718-3] [Medline: 26067402]
18. Adam F, Libier M, Oszustowicz T, Lefebvre D, Beal J, Meynadier J. Preoperative small-dose ketamine has no preemptive analgesic effect in patients undergoing total mastectomy. Anesth Analg 1999;89(2):444-447 [FREE Full text] [doi: 10.1097/00000539-199908000-00036] [Medline: 10439763]
19. Hickman SR, Mathieson KM, Bradford LM, Garman CD, Gregg RW, Lukens DW. Randomized trial of oral versus intravenous acetaminophen for postoperative pain control. Am J Health Syst Pharm 2018;75(6):367-375. [doi: 10.2146/ajhp170064] [Medline: 29523533]
20. Bryan C, Boren SA. The use and effectiveness of electronic clinical decision support tools in the ambulatory/primary care setting: a systematic review of the literature. Inform Prim Care 2008;16(2):79-91 [FREE Full text] [doi: 10.14236/jhi.v16i2.679] [Medline: 18713524]
21. Hoorn CJGM, Crijns HJGM, Dierick-van Daele ATM, Dekker LRC. Review on factors influencing physician guideline adherence in cardiology. Cardiol Rev 2019;27(2):80-86. [doi: 10.1097/CRD.0000000000000207] [Medline: 29634492]
22. Nair BG, Gabel E, Hofer I, Schwid H, Cannesson M. Intraoperative clinical decision support for anesthesia: a narrative review of available systems. Anesth Analg 2017;124(2):603-617 [FREE Full text] [doi: 10.1213/ANE.0000000000001636] [Medline: 28099325]

Abbreviations

ASA: American Society of Anesthesiology
CDSS: clinical decision support system
DSS: decision support system
ERAS: enhanced recovery after surgery
LOS: length of stay
PACU: postanesthesia care unit

XSL•FO
RenderX

PONV: postoperative nausea and vomiting

Edited by T de Azevedo Cardoso; submitted 07.11.22; peer-reviewed by U Alexis, G Lubinsky, A Dedeilia, O Shairzay, R Go; comments to author 11.05.23; revised version received 12.06.23; accepted 18.08.23; published 03.11.23.

Please cite as:
Gopwani S, Bahrun E, Singh T, Popovsky D, Cramer J, Geng X
Efficacy of Electronic Reminders in Increasing the Enhanced Recovery After Surgery Protocol Use During Major Breast Surgery: Prospective Cohort Study
JMIR Perioper Med 2023;6:e44139
URL: https://periop.jmir.org/2023/1/e44139
doi:10.2196/44139
PMID:37921854

©Sumeet Gopwani, Ehab Bahrun, Tanvee Singh, Daniel Popovsky, Joseph Cramer, Xue Geng. Originally published in JMIR Perioperative Medicine (http://periop.jmir.org), 03.11.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Perioperative Medicine, is properly cited. The complete bibliographic information, a link to the original publication on http://periop.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Real-Time Mobile Intervention to Reduce Sedentary Behavior Before and After Cancer Surgery: Pilot Randomized Controlled Trial

Carissa A Low[1], PhD; Michaela Danko[1], BSN; Krina C Durica[1], MA; Julio Vega[1], PhD; Meng Li[1], MS; Abhineeth Reddy Kunta[1], MS; Raghu Mulukutla[2], MS; Yiyi Ren[2], MS; Susan M Sereika[1], PhD; David L Bartlett[3], MD; Dana H Bovbjerg[1], PhD; Anind K Dey[2], PhD; John M Jakicic[4], PhD

[1]University of Pittsburgh, Pittsburgh, PA, United States
[2]University of Washington, Seattle, WA, United States
[3]Allegheny Health Network Cancer Institute, Pittsburgh, PA, United States
[4]University of Kansas Medical Center, Kansas City, KS, United States

Corresponding Author:
Carissa A Low, PhD
University of Pittsburgh
3347 Forbes Avenue
Suite 200
Pittsburgh, PA, 15213
United States
Phone: 1 4126235973
Email: lowca@upmc.edu

Abstract

Background: Sedentary behavior (SB) is prevalent after abdominal cancer surgery, and interventions targeting perioperative SB could improve postoperative recovery and outcomes. We conducted a pilot study to evaluate the feasibility and preliminary effects of a real-time mobile intervention that detects and disrupts prolonged SB before and after cancer surgery, relative to a monitoring-only control condition.

Objective: Our aim was to evaluate the feasibility and preliminary effects of a perioperative SB intervention on objective activity behavior, patient-reported quality of life and symptoms, and 30-day readmissions.

Methods: Patients scheduled for surgery for metastatic gastrointestinal cancer (n=26) were enrolled and randomized to receive either the SB intervention or activity monitoring only. Both groups used a Fitbit smartwatch and companion smartphone app to rate daily symptoms and collect continuous objective activity behavior data starting from at least 10 days before surgery through 30 days post discharge. Participants in the intervention group also received prompts to walk after any SB bout that exceeded a prespecified threshold, with less frequent prompts on days that patients reported more severe symptoms. Participants completed end-of-study ratings of acceptability, and we also examined adherence to assessments and to walking prompts. In addition, we examined effects of the intervention on objective SB and step counts, patient-reported quality of life and depressive and physical symptoms, as well as readmissions.

Results: Accrual (74%), retention (88%), and acceptability ratings (mean overall satisfaction 88.5/100, SD 9.1) were relatively high. However, adherence to assessments and engagement with the SB intervention decreased significantly after surgery and did not recover to preoperative levels after postoperative discharge. All participants exhibited significant increases in SB and symptoms and decreases in steps and quality of life after surgery, and participants randomized to the SB intervention unexpectedly had longer maximum SB bouts relative to the control group. No significant benefits of the intervention with regard to activity, quality of life, symptoms, or readmission were observed.

Conclusions: Perioperative patients with metastatic gastrointestinal cancer were interested in a real-time SB intervention and rated the intervention as highly acceptable, but engagement with the intervention and with daily symptom and activity monitoring decreased significantly after surgery. There were no significant effects of the intervention on step counts, patient-reported quality of life or symptoms, and postoperative readmissions, and there was an apparent adverse effect on maximum SB. Results highlight

XSL•FO
RenderX

the need for additional work to modify the intervention to make reducing SB and engaging with mobile health technology after abdominal cancer surgery more feasible and beneficial.

Trial Registration: ClinicalTrials.gov NCT03211806; https://tinyurl.com/3napwkkt

(JMIR Perioper Med 2023;6:e41425) doi:10.2196/41425

KEYWORDS

sedentary behavior; mobile health; smartphone; mobile phone; wearable device; surgical oncology; physical activity; perioperative cancer patients; surgical recovery; abdominal cancer surgery; perioperative intervention; activity monitoring

Introduction

Surgical treatment is a critical component of curative therapy for most cancers, but risks for postoperative complications, unplanned readmissions, and persistent functional impairments are common, especially for abdominal cancers, where rates of complications and readmissions can range from 25%-50% [1-3]. These high rates of adverse postoperative outcomes place patients at risk for functional limitations and impaired quality of life as well as high health care costs and utilization. Supportive interventions aimed at optimizing perioperative health and functioning are needed for this high-risk surgical oncology population.

Physical activity is one modifiable behavior that holds promise for affecting postoperative recovery and outcomes [4-6]. Indeed, prehabilitation programs that promote physical activity before surgery have been linked to improved preoperative functional capacity [7] and shorter length of stay after cancer surgery [8]. Similarly, early mobilization after surgery, generally defined as out of bed activity by postoperative day one, is recommended as part of Enhanced Recovery after Surgery pathways, although evidence of benefit is mixed [9]. Because both prehabilitation and Enhanced Recovery after Surgery often include nutritional interventions and other components, it is difficult to determine whether and to what extent increased physical activity alone can reduce postoperative risks. Moreover, postoperative symptoms such as pain and fatigue make increasing physical activity after cancer surgery challenging and may compromise adherence to exercise interventions [10,11]. In the perioperative surgical oncology context, interventions aimed at disrupting prolonged sedentary behavior (SB) with brief walking breaks may be more attainable than more structured exercise interventions, especially if the intervention can adapt to changing symptom burden over the perioperative course. To date, no studies have tested the impact of perioperative SB disruption on surgical oncology outcomes [12].

The goal of this study was to pilot-test a personalized mobile technology–supported intervention to reduce SB before and after gastrointestinal cancer surgery. This intervention uses a smartwatch and smartphone to collect daily symptom ratings that are used to tailor the frequency of prompts to disrupt prolonged SB in real time, which we hypothesized would increase the feasibility of the intervention. We previously described the development and usability of this intervention in a single-arm pilot trial [13]. In this pilot randomized controlled trial, we compared patients randomized to receive the SB intervention to those whose activity and symptoms were monitored only. The primary outcome of this initial pilot trial was feasibility, defined as accrual and retention, end-of-study acceptability ratings, and adherence to intervention assessments and activity prompts. Secondary outcomes included objective activity and SB, patient-reported quality of life and symptoms, as well as postoperative readmissions.

Methods

Participants

Participants were recruited from the outpatient clinics of 6 surgeons specializing in abdominal cancer surgery at a National Cancer Institute–designated comprehensive cancer center. Participants were recruited between June 2019 and March 2021 at their preoperative surgical oncology clinic visit. Study accrual was paused from March to December of 2020 due to the COVID-19 pandemic. Research staff provided study information and email reminders to the 6 surgical oncology care teams and asked the nurse or physician assistant to identify potential patients at the time of their consent to surgery, confirm their eligibility, and to either consent them directly or connect them with the research team for consent and onboarding. The study was open to English-speaking adults scheduled for surgical treatment of metastatic gastrointestinal or peritoneal cancer and able to stand and walk unassisted. Exclusion criteria included having less than 10 days to scheduled surgery date, to provide adequate time for participants to become familiar with study technology and activity prompts prior to surgery. No participants had sensory or motor impairments that interfered with use of the study apps.

Study Procedures

Following completion of written informed consent, participants were randomized via random number generator to either the SB intervention (which included activity monitoring) or activity monitoring only. They were provided with a Fitbit Versa smartwatch (first generation) paired with a Google Pixel 2 smartphone on which Detecting Activity to Support Healing (DASH) study apps (Intervention or Monitoring-only) as well as the Fitbit app had been installed. From the time of consent to 30 days after hospital discharge following their surgery, participants were asked to keep the devices charged, to wear the smartwatch as much as possible, to rate their daily experience of symptom severity once each morning, and for intervention participants only, to respond to activity prompts. Participants completed a questionnaire at study entry to collect information about demographic variables, health behaviors, and experience with mobile technology. Before surgery, during inpatient recovery, and approximately 30 days after postoperative discharge, participants also completed

XSL•FO
RenderX

standardized measures of depressive and physical symptoms and quality of life. At the end of the study, all participants completed a semistructured interview about their experiences with the devices and a questionnaire about the acceptability and usability of the apps.

As previously described [13], all participants used the DASH Android smartphone study app to rate the daily severity of 10 symptoms (ie, pain, fatigue or tiredness, sleep disturbance, trouble concentrating or remembering things, feeling sad or down, feeling anxious or worried, shortness of breath, numbness or tingling, nausea, and diarrhea or constipation), using a scale from 0 (ie, symptom not present) to 10 (ie, symptom as bad as you can imagine). Participants were randomized to either the DASH intervention or monitoring-only control condition. Participants randomized to the DASH intervention received a Fitbit smartwatch app that used the most recent symptom rating to set a threshold for SB bouts and used real-time step count data to trigger activity prompt notifications when prespecified SB thresholds were exceeded (60 consecutive minutes of SB when all symptoms were rated less than 7 out of 10 or 120 consecutive minutes of SB if any symptom was rated as 7 or higher). For the purposes of this study, SB was operationalized as a minute with fewer than 10 steps logged by the Fitbit, to allow for incidental stepping and arm movements that might be misclassified as steps, while also classifying very slow walking as activity, given the perioperative context and likely diminished gait cadence during early postoperative recovery [14]. When SB thresholds were exceeded, an activity prompt ("Ready for a short walk?") was sent to the smartwatch. If 30 or more steps were logged within 15 minutes of an activity prompt, participants received a positive feedback message ("Great job being active!"). Prompts were sent only between each participant's waking time and bedtime, which were set by participants in the Android app and could be adjusted during the study. Participants randomized to monitoring-only received a Fitbit smartwatch app that measured steps but did not send activity prompts.

Ethics Approval

All procedures were approved by the University of Pittsburgh Institutional Review Board (STUDY 19030389) and registered on ClinicalTrials.gov (NCT03211806).

Measures

Primary outcome measures assessing feasibility were (1) accrual and retention rates (ie, percentage of participants approached who enrolled in the research and percentage of participants enrolled who completed the study); (2) acceptability, based on end-of-study responses to the System Usability Scale [15] and the following questions: "On a scale of 0-100, how easy was it to use the smartphone/watch?" "On a scale of 0-100, how pleasant was the smartphone/watch interface (appearance, design, usability)?" and "On a scale of 0-100, how satisfied were you with the overall system (including the smartphone and watch and all notifications)?"; as well as (3) adherence (percentage of days symptom ratings were completed and at least 8 hours of Fitbit data were logged, and for intervention participants only, percentage of activity prompts after which steps were detected).

Secondary outcome measures included (1) objective SB (maximum and mean SB bout duration per day based on Fitbit minutes with less than 10 steps logged); (2) objective physical activity (Fitbit step count per day); (3) patient-reported symptoms (depressive symptoms via Center for Epidemiological Studies-Depression [16]) and physical symptoms via questions adapted from the MD Anderson Symptom Inventory and based on the National Cancer Institute's Symptom Management and Health-Related Quality of Life Steering Committee recommendations [17,18]; (4) patient-reported quality of life via the Functional Assessment of Cancer Therapy [19]; and (5) readmissions within 30 days after index hospital discharge, extracted from electronic medical records.

Analytic Approach

Group differences in baseline participant characteristics and end-of-study acceptance and usability ratings were examined using independent sample two-tailed *t* tests and chi-squared tests. Linear mixed modeling assuming the best-fitting variance-covariance structure for the repeated assessments was used to explore the effect of the intervention over the 3 study time points (ie, preoperative, inpatient, and after discharge) for the outcomes of adherence, SB, physical activity, psychological and physical symptoms, and quality of life. The models included a fixed, between-subjects effect for randomized group assignment as well as a fixed, within-subject effect for time and group interaction by time. In addition, to test statistics (F test values) and corresponding P values from the model, least square means with standard errors are presented. Two outcomes, average SB and steps, were square-root transformed due to positively skewed residual distributions when modeling the outcome in its original metric. Data for Fitbit step counts and SB bout duration were only included from days that the Fitbit was worn at least 8 hours, and sleep episodes as identified by the Fitbit were excluded from SB bouts.

Results

Participant Characteristics

As shown in Table 1, the sample was primarily White and predominantly male, with most patients undergoing cytoreductive surgery with hyperthermic intraperitoneal chemotherapy. Participants randomized to the intervention arm had significantly higher BMI and were less likely to be a former smoker compared to those randomized to the control arm. Participants started using the DASH apps a mean of 19.6 (range 8-47) days prior to surgery, throughout their inpatient stay as feasible (which lasted an average of 10.9, range 5-24 days), and for 30 days post discharge, for an average of 57.2 total days (range 44-92 days) of study participation.

XSL•FO
RenderX

Table 1. Participant characteristics.

Characteristics	All participants (n=26)	Intervention arm (n=13)	Monitoring-only arm (n=13)
Age (years), mean (SD)	56.2 (10.5)	54.9 (6.5)	57.5 (13.5)
Sex, n (%)			
Female	11 (42.3)	7 (53.8)	4 (30.8)
Male	15 (57.7)	6 (46.2)	9 (69.2)
Race, n (%)			
White	24 (92.3)	13 (100)	11 (84.6)
Black	1 (3.8)	0 (0)	1 (7.7)
More than one	1 (3.8)	0 (0)	1 (7.7)
BMI, mean (SD)	27.4 (5.3)	30.2 (5.8)	24.8 (3.3)
Smoking history, n (%)			
Current smoker	1 (4)	1 (8.3)	0 (0)
Former smoker	5 (20)	0 (0)	5 (38.5)
Never a smoker	19 (76)	11 (91.7)	8 (61.5)
Exercise frequency, n (%)			
Seldom or never	6 (24)	4 (33.3)	2 (15.4)
1-2 times per week	7 (28)	5 (41.7)	2 (15.4)
3-4 times per week	10 (40)	3 (25)	7 (53.8)
>5 times per week	2 (8)	0 (0)	2 (15.4)
Has Wi-Fi at home, n (%)	22 (88)	11 (91.7)	11 (84.6)
Owns a smartphone, n (%)	26 (100)	12 (100)	13 (100)
Owns an activity tracker, n (%)	4 (16)	2 (16.7)	2 (15.4)
CS+HIPEC[a] surgery, n (%)	17 (65.4)	10 (76.9)	7 (53.8)

[a]CS+HIPEC: cytoreductive surgery with hyperthermic intraperitoneal chemotherapy.

Primary Outcomes

Accrual and Retention

Of the 35 eligible patients approached, 26 consented to the study (74% accrual rate). Reasons for not participating were "too busy/overwhelmed" (n=2), "not good with technology" (n=3), "already wear a smartwatch/activity monitor and did not want to wear two" (n=2), and "had to leave clinic so did not have time to discuss the study" (n=2). The retention rate for the study was 88%, with 3 participants withdrawing (2 participants before starting to use the devices—one in intervention and one in monitoring-only condition—and 1 participant in the intervention condition 18 days after surgery due to poor health and readmission).

Acceptance

A total of 20 participants completed the end-of-study interview, and those in both the intervention and monitoring-only conditions rated the phone and watch interfaces as pleasant and easy to use and the overall system as satisfactory and usable (Table 2).

Table 2. Mean (SD) participant ratings of interface and system usability.

Variable (range 0-100)	All (n=20)	Intervention arm (n=9)	Monitoring-only arm (n=11)	*P* value
Phone—ease of use	93.1 (7.2)	91.1 (8.3)	94.6 (6.1)	.29
Watch—ease of use	91.6 (12.6)	94.4 (5.1)	89.3 (16.3)	.38
Phone—pleasantness	87.7 (13.7)	90.0 (13.2)	85.8 (14.5)	.51
Watch—pleasantness	87.9 (14.9)	91.7 (7.9)	84.8 (18.7)	.32
Overall satisfaction	88.5 (9.1)	91.3 (5.5)	86.2 (11.0)	.22
System Usability Scale	85.1 (11.5)	83.6 (9.2)	86.4 (13.5)	.61

When asked what they thought of the study, participants in the intervention condition reported that the activity prompts were motivating, especially prior to surgery. For example, one participant (P13) noted that "it got me moving more than I

XSL•FO
RenderX

normally would have." However, participants also noted that the prompts were not as motivating or as easy to respond to after surgery, especially in the hospital. One participant (P6) said the following:

> *At the beginning, I thought it was awesome and was very excited about the step counting and found it motivational; after surgery, I had a lot of trouble getting interest back, and the watch wasn't enough to be motivating; I lost interest because I had other priorities health-wise.*

Another participant (P7) said the following:

> *It's so much easier to get up presurgery. Maybe a hierarchy of prompts tapping different motivations [would be better] because it takes so much more to get up post-surgery.*

Across both conditions, participants mentioned that they enjoyed tracking their step or sleep data in the Fitbit app. One participant (P16, in the monitoring-only group) noted the following:

> *Part of my recovery was setting step goals for myself and increasing that goal.*

Some participants, like P2 (in the monitoring-only group), also noticed associations between activity and how they felt, noting "days with higher steps always felt better symptom-wise, looking back." Many participants felt that physical activity was beneficial for their physical and psychological recovery, as P13 (in the intervention group) said the following:

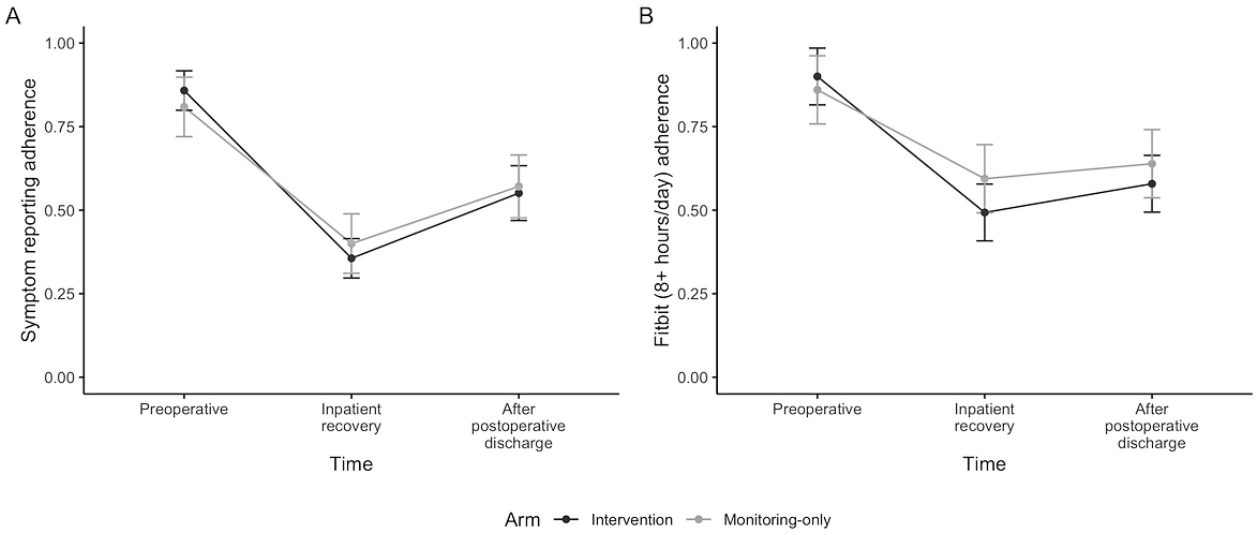
> *Moving and walking helped prevent scar tissue development. If I had stayed sedentary, I think I would have been in much worse shape.*

Adherence

Over the course of the study, daily symptom ratings were completed on 62% (874/1416) of days, ranging from 14% (9/65) to 95% (55/58) of days across individual participants. Fitbits were worn on 77% (1091/1416) of study days, and 91% (990/1091) of these days had at least 8 hours of Fitbit data available. On average, 69% (977/1416) of days were included in Fitbit analyses; across individual participants, the percentage of days with at least 8 hours of Fitbit data ranged from 17% (9/52 days with ≥8 hours of data) to 100% (58/58 days with ≥8 hours of data). As shown in Figure 1, participants became less adherent with both symptom reporting and wearing the Fitbit after surgery (symptom reporting: F_{time}=22.9; P<.001; and Fitbit: F_{time}=9.2; P=.001), but there were no significant differences in adherence between the two study groups (symptom reporting: F_{group}=0.0; P=.95; $F_{group\times time}$=0.3; P=.78; and Fitbit: F_{group}=0.2; P=.663; $F_{group\times time}$=0.4; P=.68).

For participants in the intervention group, an average of 5.8 activity prompts were sent per day, and participants took steps and received positive feedback after 22% (418/1925) of activity prompts. This varied substantially from before surgery (mean 3.3, SD 1.8 prompts per day; 200/407, 49% of prompts resulted in walking) to after surgery in the hospital (mean 7.8, SD 2.6 prompts per day; 29/462, 6% of prompts resulted in walking) to postdischarge recovery (mean 6.2, SD 2.6 prompts per day; 189/1056, 18% of prompts resulted in walking).

Figure 1. Proportion of days (in mean and SE) participants were adherent with (A) daily symptom reporting and (B) wearing the Fitbit for at least 8 hours per day.



Secondary Outcomes

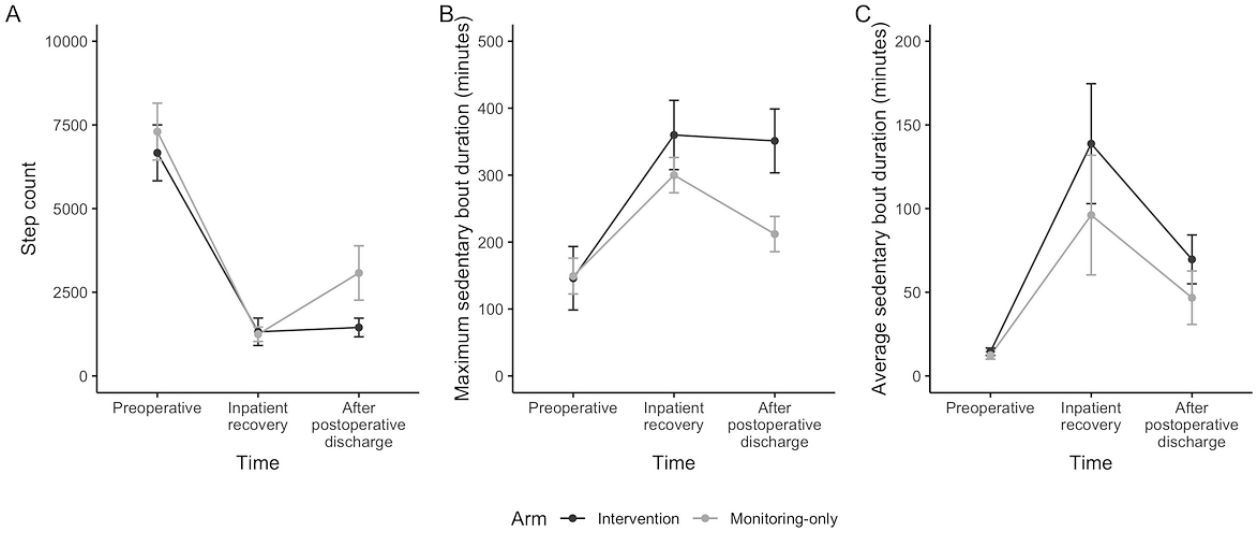
Fitbit-Measured Sedentary Behavior Bouts and Steps

On average, participants logged 3642 (SD 3365) steps per day with a mean SB bout duration of 61 (SD 80) minutes and a maximum SB bout duration of 248 (SD 155) minutes. For all participants, step counts decreased significantly, and mean and maximum SB bout duration increased significantly from before surgery to during inpatient recovery (Figure 2; step count: F_{time}=60.5; P<.001; maximum SB bout: F_{time}=10.1; P<.001; and mean SB bout: F_{time}=28.1; P<.001). The intervention had no significant effect on step count (F_{group}=2.3; P=.15; and $F_{group\times time}$=1.2; P=.32) or mean SB bout duration (F_{group}=1.5; P=.24; and $F_{group\times time}$=0.5; P=.24). Unexpectedly, participants randomized to the intervention had *longer* maximum SB bouts overall (F_{group}=6.16; P=.02; and $F_{group\times time}$=1.48; P=.24). One

XSL•FO
RenderX

important limitation to note is that these mean step count and SB bout values are based on the subset of participants who were compliant with wearing the smartwatch. Although there were no significant group differences in Fitbit compliance, intervention participants tended to wear the watch for fewer hours per day, and some mentioned removing the watch when they knew they would not be able to get up and walk or when trying to nap or rest. Because we included all days with at least 8 hours of total but not necessarily consecutive wear time, we may have misclassified some episodes during which participants were not wearing the watch as sedentary bouts. When hours of wear time per day was included in models, the group effect on maximum SB bout duration was no longer significant (data not shown).

Figure 2. (A) Fitbit daily mean step count, (B) maximum sedentary bout duration, and (C) mean sedentary bout duration.

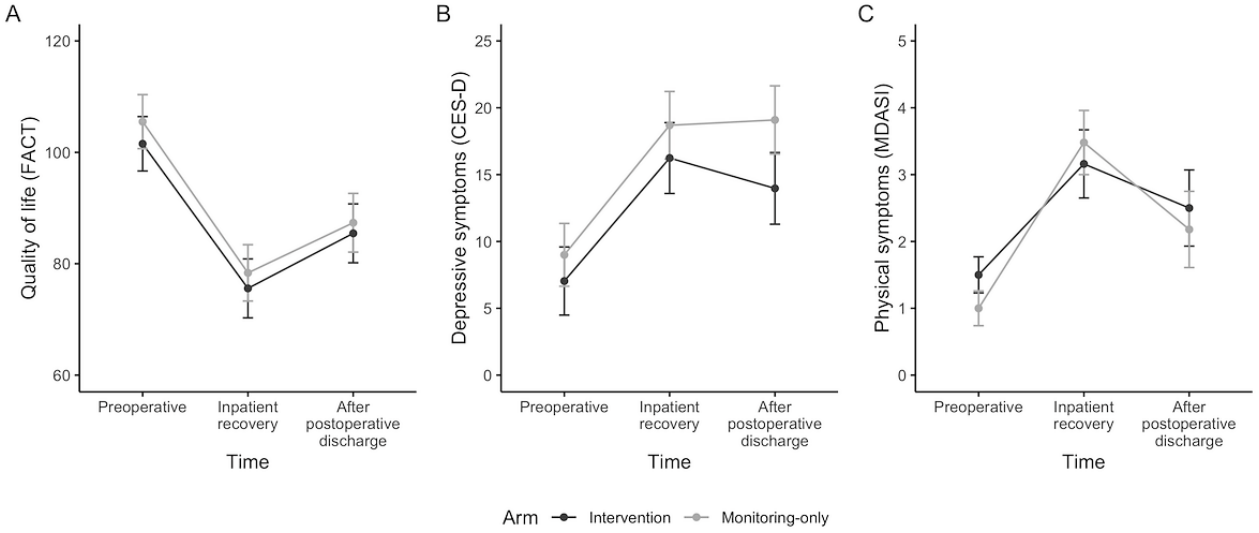


Patient-Reported Measures

Similar to the other outcomes, we observed a significant time effect for quality of life (F_{time}=21.4; P<.001; Figure 3), depressive symptoms (F_{time}=10.9; P<.001), and physical symptoms (F_{time}=24.0; P<.001), but no significant group (quality of life: F_{group}=0.3; P=.60; depressive symptoms: F_{group}=1.6; P=.22; and physical symptoms: F_{group}=0.1; P=.76) or group × time effect (quality of life: $F_{group\times time}$=0.0; P=.97; depressive symptoms: $F_{group\times time}$=0.3; P=.78; and physical symptoms: $F_{group\times time}$=0.9; P=.41). All participants regardless of condition reported worsening quality of life and symptoms after surgery.

In total, 5 of 12 participants who started in the intervention condition were readmitted within 30 days, compared to 4 of 12 participants who started in the monitoring-only conditions, and there was no significant group difference in readmission rate (χ^1_2=0.2; P=.67).

Figure 3. (A) Patient-reported quality of life, (B) depressive symptoms, and (C) physical symptoms. CES-D: Center for Epidemiological Studies-Depression; FACT: Functional Assessment of Cancer Therapy; MDASI: MD Anderson Symptom Inventory.



XSL•FO
RenderX

Discussion

Principal Findings

In this paper, we described results from a pilot randomized trial testing a SB intervention in patients undergoing surgery for metastatic gastrointestinal cancer. To our knowledge, this is the first SB intervention developed specifically for patients undergoing cancer surgery at high risk for adverse postoperative outcomes [12]. Although patients were willing to participate and remain in the trial and rated the intervention as highly acceptable, engagement with the intervention and with daily symptom and activity monitoring decreased significantly after surgery. There were no significant effects of the intervention on step counts, patient-reported quality of life or symptoms, and postoperative readmissions. Contrary to hypotheses, participants randomized to the intervention group exhibited longer maximum SB bouts relative to the monitoring-only condition.

The SB intervention tested in this trial was designed to make replacing prolonged SB with short walking breaks more feasible in the perioperative context by reducing the frequency of SB prompts on days that patients reported high symptom burden. Given the significant drops in adherence and engagement that occurred after surgery and particularly during inpatient recovery as well as the fact that participants may have been less likely to complete symptom ratings on days they felt particularly unwell, additional modifications to the intervention are needed to make postoperative activity more feasible. A number of participants in the intervention condition noted that it was very difficult to get out of bed to walk when prompted, especially without assistance; this challenge and the associated frustration could have led to decreased self-efficacy to adhere to the intervention that carried into the postdischarge period, leading to increased SB and decreased adherence. Pausing the intervention until patients were recovering at home, adapting the intervention to involve caregivers or hospital staff and timing prompts around their availability to assist with postoperative ambulation, or replacing walking with light stretches or activities that could be done in bed while seated could all be options for future interventions targeting activity among postoperative inpatients. Adherence to assessments also decreased in the monitoring-only group after surgery, suggesting that either reduced frequency of assessments or additional support and reminders may be needed to make collection of continuous activity and daily symptom ratings feasible postoperatively.

The lack of observed benefits with regard to SB and activity may have been related to low adherence and engagement, or the intervention may not have been sufficiently robust to produce a change in activity during the acute perioperative period. Although participants in our usability and feasibility study reported that the frequency of prompts was appropriate, the current intervention was fairly minimal, aiming to disrupt SB bouts of 1-2 hours with a small (30 steps or more) amount of walking and positive reinforcement when goals were met. Participants may have disengaged from the intervention if they did not perceive it to be beneficial, and targeting a higher activity goal, if done in a feasible way that factors in physical limitations after surgery, could result in higher adherence over time if participants perceive the SB intervention to be meaningfully increasing activity and to have potential health benefits. In the future, providing education about the risks of perioperative SB, personalized goal setting to inform prompts, and coaching and problem-solving to overcome barriers to disrupting SB could be considered as additional interventional components to enhance a perioperative SB intervention [12,20]. Involving patients in the co-design of perioperative SB interventions could also result in enhanced engagement and benefit [21].

Comparison to Prior Work

Although adherence to reporting symptoms and wearing the Fitbit declined significantly after surgery, rates were consistent with other work on wearables [22] and symptom reporting [23] during cancer treatment and with an earlier study of activity monitoring after cancer surgery [24]. As in our earlier field trial [13], adherence also varied substantially between participants, ranging from approximately 15% to 100% for both symptom reporting and Fitbit wearing throughout the perioperative period. Another approach to consider in future work is a more stepped-care approach, with more frequent contact or high-touch support for patients with poor adherence. Of note, because we used the same wearable device to both deliver the intervention and measure objective activity, poor adherence and engagement resulted in less accurate assessments of activity and SB, which may also have affected results, particularly if patients in the intervention group began wearing the device for fewer hours each day after surgery due to inability to respond to the walking prompts or to minimize disruptions caused by the prompts.

Although the intervention yielded no significant benefits for patients, this study highlights the continued need for interventions to improve postoperative recovery following surgery for metastatic abdominal cancer. Consistent with other studies, nearly 40% of patients in our sample experienced an unplanned hospital readmission, and participants remained significantly less active 30 days after postoperative discharge, relative to their preoperative activity levels. In addition to improved interventions targeting SB and activity, interventions aimed at remotely monitoring and addressing worsening pain or other symptoms as well as other causes of readmission (eg, dehydration) hold promise for their ability to support this high-risk population.

Strengths and Limitations

Strengths of this study include the randomized design and the use of real-time symptom ratings and step data to trigger personalized just-in-time activity prompts. Focusing on disrupting SB rather than increasing physical activity is novel in the context of cancer surgery. Starting the intervention prior to scheduled surgery allowed participants to become familiar with the devices and begin increasing activity prior to surgery and hospitalization, while there may also be clinical value in continuing an intervention shortly after surgery when SB is very prevalent. The use of off-the-shelf consumer devices is also highly scalable.

XSL•FO
RenderX

This study had a number of important limitations. First, the sample size was smaller than originally intended due to COVID-19 pandemic–related disruptions to accrual and is an important limitation of this study. Second, developing a system capable of remotely detecting real-time step counts using a consumer wearable device proved challenging and required us to provide study Android phones to participants to use for collecting symptom ratings and synchronizing the Fitbit smartwatch. All enrolled participants already owned a personal smartphone; requiring them to also carry and charge an additional study smartphone across perioperative transitions of care may have contributed to adherence challenges; future interventions in this area should be deployed on participants' existing phones to potentially improve feasibility. Third, we elected to use a monitoring-only control so that the only difference between the two study groups was the activity prompts in recognition of the fact that merely using an activity monitor can promote physical activity among cancer patients [25]; alternative control conditions could have yielded different results. Finally, all participants were undergoing surgery for metastatic peritoneal or gastrointestinal cancer, and results may not generalize to other perioperative groups or contexts.

Future Directions

As described above, additional intervention refinement and testing is needed to make real-time SB disruption more feasible and engaging for an abdominal cancer surgery population, particularly during the postoperative period. Given the small sample in this work, larger trials of activity modification are necessary once the intervention has been improved to be more feasible and potentially more robust. There have been significant advances in consumer wearable technology since the DASH apps were developed in 2018, and future work should consider interventions that leverage Apple HealthKit or GoogleFit and work across different activity monitoring devices.

Conclusions

In conclusion, although patients undergoing abdominal cancer surgery were interested in a real-time SB intervention and rated the intervention as highly acceptable, adherence and engagement decreased significantly after surgery, and there were no observed benefits of the intervention on objective activity, quality of life, symptoms, or readmissions. Further research may be needed to understand factors that influence SB following surgery and to make reducing SB and engaging with mobile health technology after surgery more feasible and beneficial for these high-risk patients.

Acknowledgments

The authors gratefully acknowledge the assistance of Michelle Liebdzinksi and Heather Jones in recruiting patients as well as the patient participants who made this research possible. This research was funded by the National Cancer Institute (K07CA204380 and P30CA047904).

Data Availability

Data generated and analyzed during this study are not publicly available to protect patient health information but are available as deidentified data from the corresponding author on reasonable request.

Conflicts of Interest

JMJ is part of the Scientific Advisorty Board for Wondr Health, Inc, is a Consultant for Educational Initiatives, Inc, and is the Principal Investigator on a resesrch grant awarded to the University of Kansas Medical Center by Epitomee Medical, Inc.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.2).
[PDF File (Adobe PDF File), 95 KB - periop_v6i1e41425_app1.pdf]

References

1. Choudry MHA, Shuai Y, Jones HL, Pai RK, Pingpank JF, Ahrendt SS, et al. Postoperative complications independently predict cancer-related survival in peritoneal malignancies. Ann Surg Oncol 2018 Dec 09;25(13):3950-3959. [doi: 10.1245/s10434-018-6823-9] [Medline: 30302637]
2. Shan LL, Saxena A, Shan BL, Morris DL. Quality of life after cytoreductive surgery and hyperthermic intra-peritoneal chemotherapy for peritoneal carcinomatosis: a systematic review and meta-analysis. Surg Oncol 2014 Dec;23(4):199-210. [doi: 10.1016/j.suronc.2014.10.002] [Medline: 25466850]
3. Rieser CJ, Hoehn RS, Zenati M, Hall LB, Kang E, Zureikat AH, et al. Impact of socioeconomic status on presentation and outcomes in colorectal peritoneal metastases following cytoreduction and chemoperfusion: persistent inequalities in outcomes at a high-volume center. Ann Surg Oncol 2021 Jul 09;28(7):3522-3531 [FREE Full text] [doi: 10.1245/s10434-021-09627-2] [Medline: 33687614]
4. Carli F, Charlebois P, Stein B, Feldman L, Zavorsky G, Kim DJ, et al. Randomized clinical trial of prehabilitation in colorectal surgery. Br J Surg 2010 Aug;97(8):1187-1197. [doi: 10.1002/bjs.7102] [Medline: 20602503]

XSL•FO
RenderX

5. Singh F, Newton RU, Galvão DA, Spry N, Baker MK. A systematic review of pre-surgical exercise intervention studies with cancer patients. Surg Oncol 2013 Jun;22(2):92-104. [doi: 10.1016/j.suronc.2013.01.004] [Medline: 23434347]
6. de Almeida E, de Almeida J, Landoni G, Galas FRBG, Fukushima J, Fominskiy E, et al. Early mobilization programme improves functional capacity after major abdominal cancer surgery: a randomized controlled trial. BJA 2017 Nov;119(5):900-907. [doi: 10.1093/bja/aex250]
7. Waterland JL, McCourt O, Edbrooke L, Granger CL, Ismail H, Riedel B, et al. Efficacy of prehabilitation including exercise on postoperative outcomes following abdominal cancer surgery: a systematic review and meta-analysis. Front Surg 2021 Mar 19;8:628848 [FREE Full text] [doi: 10.3389/fsurg.2021.628848] [Medline: 33816546]
8. Lambert J, Hayes LD, Keegan TJ, Subar DA, Gaffney CJ. The impact of prehabilitation on patient outcomes in hepatobiliary, colorectal, and upper gastrointestinal cancer surgery: a PRISMA-accordant meta-analysis. Ann Surg 2021 Jul 01;274(1):70-77. [doi: 10.1097/SLA.0000000000004527] [Medline: 33201129]
9. Castelino T, Fiore JF, Niculiseanu P, Landry T, Augustin B, Feldman LS. The effect of early mobilization protocols on postoperative outcomes following abdominal and thoracic surgery: a systematic review. Surgery 2016 Apr;159(4):991-1003. [doi: 10.1016/j.surg.2015.11.029] [Medline: 26804821]
10. Clark MM, Vickers KS, Hathaway JC, Smith M, Looker SA, Petersen LR, et al. Physical activity in patients with advanced-stage cancer actively receiving chemotherapy. J Support Oncol 2007;5(10):487-493. [Medline: 18240671]
11. Courneya K, Segal RJ, Gelmon K, Reid RD, Mackey JR, Friedenreich CM, et al. Predictors of supervised exercise adherence during breast cancer chemotherapy. Med Sci Sports Exerc 2008 Jun;40(6):1180-1187. [doi: 10.1249/MSS.0b013e318168da45] [Medline: 18460985]
12. Belcher BR, Kang D, Yunker AG, Dieli-Conwright CM. Interventions to reduce sedentary behavior in cancer patients and survivors: a systematic review. Curr Oncol Rep 2022 Nov 13;24(11):1593-1605. [doi: 10.1007/s11912-022-01313-0] [Medline: 35829982]
13. Low CA, Danko M, Durica KC, Kunta AR, Mulukutla R, Ren Y, et al. A real-time mobile intervention to reduce sedentary behavior before and after cancer surgery: usability and feasibility study. JMIR Perioper Med 2020 Mar 23;3(1):e17292 [FREE Full text] [doi: 10.2196/17292] [Medline: 33393915]
14. Tudor-Locke C, Han H, Aguiar EJ, Barreira TV, Schuna JM, Kang M, et al. How fast is fast enough? Walking cadence (steps/min) as a practical estimate of intensity in adults: a narrative review. Br J Sports Med 2018 Jun;52(12):776-788 [FREE Full text] [doi: 10.1136/bjsports-2017-097628] [Medline: 29858465]
15. Brooke J. System usability scale (SUS): a quick-and-dirty usability scale. In: Usability Evaluation In Industry. Boca Raton, FL: CRC Press; 1996.
16. Radloff LS. The CES-D Scale: a self-report depression scale for research in the general population. Appl Psychol Meas 2016 Jul 26;1(3):385-401. [doi: 10.1177/014662167700100306]
17. Cleeland CS, Mendoza TR, Wang XS, Chou C, Harle MT, Morrissey M, et al. Assessing symptom distress in cancer patients. Cancer 2000 Oct 01;89(7):1634-1646. [doi: 10.1002/1097-0142(20001001)89:7<1634::aid-cncr29>3.0.co;2-v]
18. Reeve BB, Mitchell SA, Dueck AC, Basch E, Cella D, Reilly CM, et al. Recommended patient-reported core set of symptoms to measure in adult cancer treatment trials. J Natl Cancer Inst 2014 Jul 08;106(7):dju129-dju129 [FREE Full text] [doi: 10.1093/jnci/dju129] [Medline: 25006191]
19. Cella DF, Tulsky DS, Gray G, Sarafian B, Linn E, Bonomi A, et al. The Functional Assessment of Cancer Therapy scale: development and validation of the general measure. J Clin Oncol 1993 Mar;11(3):570-579. [doi: 10.1200/JCO.1993.11.3.570] [Medline: 8445433]
20. Lam K, Baurecht H, Pahmeier K, Niemann A, Romberg C, Biermann-Stallwitz J, et al. How effective and how expensive are interventions to reduce sedentary behavior? An umbrella review and meta-analysis. Obes Rev 2022 May 23;23(5):e13422. [doi: 10.1111/obr.13422] [Medline: 35068047]
21. Robinson A, Oksuz U, Slight R, Slight S, Husband A. Digital and mobile technologies to promote physical health behavior change and provide psychological support for patients undergoing elective surgery: meta-ethnography and systematic review. JMIR Mhealth Uhealth 2020 Dec 01;8(12):e19237 [FREE Full text] [doi: 10.2196/19237] [Medline: 33258787]
22. Beauchamp UL, Pappot H, Holländer-Mieritz C. The use of wearables in clinical trials during cancer treatment: systematic review. JMIR Mhealth Uhealth 2020 Nov 11;8(11):e22006 [FREE Full text] [doi: 10.2196/22006] [Medline: 33174852]
23. Warrington L, Absolom K, Conner M, Kellar I, Clayton B, Ayres M, et al. Electronic systems for patients to report and manage side effects of cancer treatment: systematic review. J Med Internet Res 2019 Jan 24;21(1):e10875 [FREE Full text] [doi: 10.2196/10875] [Medline: 30679145]
24. de Leeuwerk ME, Botjes M, van Vliet V, Geleijn E, de Groot V, van Wegen E, et al. Self-monitoring of physical activity after hospital discharge in patients who have undergone gastrointestinal or lung cancer surgery: mixed methods feasibility study. JMIR Cancer 2022 Jun 24;8(2):e35694 [FREE Full text] [doi: 10.2196/35694] [Medline: 35749165]
25. Singh B, Zopf EM, Howden EJ. Effect and feasibility of wearable physical activity trackers and pedometers for increasing physical activity and improving health outcomes in cancer survivors: a systematic review and meta-analysis. J Sport Health Sci 2022 Mar;11(2):184-193 [FREE Full text] [doi: 10.1016/j.jshs.2021.07.008] [Medline: 34314878]

XSL•FO
RenderX

Abbreviations

DASH: Detecting Activity to Support Healing
SB: sedentary behavior

Edited by T Leung; submitted 26.07.22; peer-reviewed by KS Sahu, YC Wang, M Kapsetaki; comments to author 25.11.22; revised version received 16.12.22; accepted 23.12.22; published 12.01.23.

Please cite as:
Low CA, Danko M, Durica KC, Vega J, Li M, Kunta AR, Mulukutla R, Ren Y, Sereika SM, Bartlett DL, Bovbjerg DH, Dey AK, Jakicic JM
A Real-Time Mobile Intervention to Reduce Sedentary Behavior Before and After Cancer Surgery: Pilot Randomized Controlled Trial
JMIR Perioper Med 2023;6:e41425
URL: https://periop.jmir.org/2023/1/e41425
doi:10.2196/41425
PMID:36633893

©Carissa A Low, Michaela Danko, Krina C Durica, Julio Vega, Meng Li, Abhineeth Reddy Kunta, Raghu Mulukutla, Yiyi Ren, Susan M Sereika, David L Bartlett, Dana H Bovbjerg, Anind K Dey, John M Jakicic. Originally published in JMIR Perioperative Medicine (http://periop.jmir.org), 12.01.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Perioperative Medicine, is properly cited. The complete bibliographic information, a link to the original publication on http://periop.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Mobile App for Postoperative Pain Management Among Older Veterans Undergoing Total Knee Arthroplasty: Mixed Methods Feasibility and Acceptability Pilot Study

Jessica Kelley Morgan[1], PhD; Caitlin R Rawlins[2], RN, MSN; Steven K Walther[1], MBA; Andrew Harvey[1], BS; Annmarie O'Donnell[1], BS; Marla Greene[2], RN, BSN; Troy G Schmidt[2], MD

[1]Continuous Precision Medicine, Research Triangle Park, NC, United States

[2]Western North Carolina Veterans Affairs Health Care System, Asheville, NC, United States

Corresponding Author:
Jessica Kelley Morgan, PhD
Continuous Precision Medicine
800 Park Offices Drive
Suite 3718
Research Triangle Park, NC, 27709-3547
United States
Phone: 1 9196185405
Email: jmorgan@cpmed.io

Abstract

Background: Prescription opioid misuse risk is disproportionate among veterans; military veterans wounded in combat misuse prescription opioids at an even higher rate (46.2%). Opioid misuse is costly in terms of morbidity, mortality, and humanitarian and economic burden and costs the Civilian Health and Medical Program of the Department of Veterans Affairs more than US $1.13 billion annually. Preventing opioid misuse at the time of prescription is a critical component in the response to the opioid crisis. The CPMRx mobile app has been shown to decrease the odds of opioid misuse during the postoperative period.

Objective: The overarching purpose of this feasibility pilot study was to explore whether deploying a mobile app (CPMRx) to track postoperative pain and medication use is feasible in a Department of Veterans Affairs medical center. In support of this goal, we had four complementary specific aims: (1) determine the technological and logistical feasibility of the mobile app, (2) assess the acceptability of the mobile app to participants, (3) measure demand for and engagement with the mobile app, and (4) explore the potential use of the mobile app to patients and providers.

Methods: Participants (N=10) were veterans undergoing total knee arthroplasty within the Veterans Health Administration provided with the CPMRx app to self-manage their pain during their 7-day at-home recovery following surgery. CPMRx uses scientifically validated tools to help clinicians understand how a patient can use the least amount of medication while getting the most benefit. The suite of software includes a mobile app for patients that includes a behavioral health intervention and a clinical decision support tool for health care providers that provides feedback about pain and medication use trends. Patients filled out paper questionnaires regarding acceptability at their postoperative follow-up appointment.

Results: Overall, quantitative measures of acceptability were high. The average rating for the amount of time required to use the app was 4.9 of 5 (5=“very little”), and the average rating for ease of use was 4.4 of 5 (5=“very easy”). Open-ended questions also revealed that most participants found ease of use to be high. Demand and engagement were high as well with a mean number of mobile app entries of 34.1 (SD 20.1) during the postoperative period. There were no reported technological or logistical issues with the mobile app. Participants took an average of 25.13 (SD 14.37) opioid tablets to manage their postoperative pain.

Conclusions: Results of this study revealed that the use of a mobile app for pain and medication management during postoperative recovery was both feasible and acceptable in older veterans undergoing total knee arthroplasty within the Veterans Health Administration. The wide variation in opioid consumption across participants revealed the potential use of the mobile app to provide actionable insights to clinicians if adopted more widely.

(JMIR Perioper Med 2023;6:e50116) doi:10.2196/50116

XSL•FO
RenderX

KEYWORDS

mobile app; pain management; opioids; older adults; veterans; mobile phone

Introduction

Background

The opioid crisis has resulted in a significant increase in opioid misuse and opioid use disorder; as a result, opioid-related overdose deaths have climbed to staggeringly high levels. Opioid misuse, use disorders, and overdoses are costly in terms of morbidity, mortality, and humanitarian and economic burden. It is estimated that in 2019 alone, more than 9.7 million adults living in the United States (3.5%) misused prescription opioids [1], and drug overdose deaths in the United States rose 29.4% in 2020 to an estimated 93,331, including 69,710 (74.7%) involving opioids [2,3]. Prescription opioid misuse risk is disproportionate among veterans; military veterans wounded in combat misuse prescription opioids at an even higher rate (46.2%) [4].

The cost of opioid overdose, abuse, and dependence in the United States is estimated to be US $78.5 billion annually [5]. If reduced quality of life from opioid use disorder and the value of life lost due to fatal opioid overdose are included, the estimate increases to more than US $1 trillion [6]. Opioid misuse costs the Civilian Health and Medical Program of the Department of Veterans Affairs more than US $1.13 billion annually [6]. Adjusted annual health care costs for diagnosed opioid misuse patients are higher than those for patients without diagnosed misuse, and the prevalence of diagnosed opioid misuse is almost 7 times higher for those in the Veterans Health Administration than in commercial health plans, translating to a significant economic burden for this population [7].

One known risk factor for opioid misuse and dependence is opioid prescription following surgical procedures, both minor and major [8]. Extant data suggest that nonmedical prescription opioid use is a strong risk factor for heroin use initiation among both civilian [9] and veteran populations [10]. For the patient being prescribed opioids, there is a need to ensure appropriate use to prevent habit-forming behaviors during postoperative recovery, as it has been estimated that 12.5% of people who are prescribed opioids misuse them [11] and, as noted earlier, this is even more pronounced among some veterans (46.2%) [4]. For members of the patient's family and others in the community, it is important that excess prescription opioids are not made available for misuse; indeed, about half (50.8%) of people who misused prescription pain relievers reported obtaining them from a friend or relative, either for free, by purchasing them, or by taking them without asking, and an additional 35.7% got them through a prescription by a single doctor [1]. Several recent studies have also found wide variation in postoperative prescribing practices and noted systemic overprescription [12-14], which may inadvertently promote continued opioid use after their indication is no longer warranted. Taken together, these findings underscore the criticality of managing the use of opioids during the postoperative period in combating the opioid crisis and highlight this time as an essential point of intervention.

Mobile Health

Over 97% of Americans own cell phones, 85% of which are smartphones. Smartphone ownership is even more common among younger adults; 96% of persons aged between 18 and 29 years, 95% of persons aged between 30 and 49 years, and 83% of persons aged between 50 and 64 years own smartphones [15]. As a result, technology-based health interventions are becoming increasingly common. Among older adults, however, smartphone ownership is less common; among adults over the age of 65 years, smartphone ownership drops to 61% [15].

Mobile health (mHealth) includes everything from the use of mobile technology, such as smartphones, to software, including mobile apps, to facilitate or enhance health care [16]. Mobile technologies are used in multiple ways, including the use of smart apps, web-based software, and SMS text messaging [17]. Smart apps are being used to augment treatment for substance abuse disorders, promote self-management of chronic conditions, increase engagement with health research, assess or measure symptoms, and foster adherence to both treatments and appointments [18-21]. Evidence suggests that symptom reporting with smart apps is well tolerated by patients and has better validity and reliability than hard copies that often rely on recall [22]. One pilot study found that 78% of patients with chronic pain who downloaded an app to track their pain symptom ratings used it with an average of 16.4 daily assessments submitted by patients within the first month. Patients reported that the app was easy to use and were willing to continue to use the app even after the study was complete [22]. Additionally, electronic monitoring (EM) of prescription medication use is a budding technology with a variety of applications and can be integrated into most drug packaging, including pill dispensers and blister packs, to help measure adherence [23]. Prior research has shown that, among veterans receiving outpatient care for posttraumatic stress disorder, age significantly predicted ownership of mHealth devices but not use or interest in mHealth apps among device owners [16]. Therefore, although older veterans may be less likely to own a personal smart device, access should be adequate, and interest in using mHealth apps is high [16].

CPMRx App

Continuous Precision Medicine has developed a suite of software, CPMRx, that uses scientifically validated tools to help clinicians understand how a patient can use the least amount of medication while getting the most benefit. The suite of software includes a mobile app for patients that includes a behavioral health intervention and a clinical decision support tool for health care providers that provides feedback about pain and medication use trends.

We have used a public health research pipeline approach for designing, evaluating, and implementing behavioral health interventions, and the CPMRx mobile app has been shown to be feasible and acceptable for supporting postoperative pain management across several patient populations and surgery types, including young adults undergoing third molar extraction

XSL•FO
RenderX

(dental clinic, Womack Army Medical Center) [24,25], adults undergoing tonsillectomy (ear, nose, and throat, WakeMed Hospital) [26], and women undergoing cesarean section (obstetrics and gynecology, Temple University Health) [27,28]. The effectiveness of the mobile app and clinical decision support tool has also been validated. The implementation of CPMRx provides actionable insights that can be used to establish more precise prescribing guidelines. At Womack Army Medical Center, for example, clinicians were able to reduce the overprescription of an opioid by 10,000 pills annually for a single surgery type (ie, third molar extraction) [24]. In a recently conducted randomized controlled trial at Temple University (N=100), the CPMRx app was shown to reduce the odds of prescription opioid misuse during the postoperative period by 92% [28].

Specific Aims

The overarching purpose of this single-arm prospective pilot study was to explore whether deploying a mobile app to track postoperative pain and medication use is feasible in a Department of Veterans Affairs Medical Center. In support of this goal, we had four complementary specific aims: (1) determine the technological and logistical feasibility of the mobile app, (2) assess the acceptability of the mobile app to participants, (3) measure demand for and engagement with the mobile app, and (4) explore the potential use of the mobile app to patients and providers.

Methods

Participants and Procedures

Participants (N=10) were veterans undergoing total knee arthroplasty at the Charles George VA Medical Center (CGVAMC). Veterans used the CPMRx app to self-manage their pain and track medication use during their 7-day at-home recovery following surgery.

A convenience sample of participants was recruited during presurgical appointments at the CGVAMC. Presurgical appointments are conducted by orthopedic nurse case managers for patients undergoing total knee arthroplasty performed by CGVAMC orthopedic surgeons. During these appointments, patients are educated about the procedure and scheduled for a physical, history, and anesthesia evaluation to ensure they are eligible for the procedure. For this study, in addition to scheduling the physical, history, and anesthesia evaluation, the orthopedic nurse case manager assisted in recruitment by briefing the patient on the research study. The orthopedic nurse coordinator made note of all patients who expressed interest in participating in the research study. During the preoperative appointment, patients provided informed consent and HIPAA (Health Insurance Portability and Accountability Act) authorization.

On the day of discharge from the postsurgical ward, patients typically receive patient education from the registered nurse assigned to them. The discharging provider ensures an order is placed for standard postoperative medications that include opioid pain medication (oxycodone 5 mg every 6 hours as needed or hydromorphone 2 mg every 6 hours as needed, 42 in total) and acetaminophen (650 mg every 6 hours as needed, 84 tablets of 325 mg in total), as well as the other medications typically prescribed following a patient's respective procedure. For this study, all pain medications provided to study participants were placed into a smart EM blister pack to record time stamps whenever a medication was removed by the patient. All participants received discharge instructions and education as part of the standard of care and additionally received instruction on the mobile app, tablet, and EM blister packs.

Participants self-managed their pain at home and used the mobile app and EM blister packs for the first week (7 days) following discharge to record pain score and medication use, as necessary. Patients met with the orthopedic nurse case manager for a typically scheduled postoperative appointment between 7 and 10 days after surgery. At this appointment, they returned the tablet and EM blister packs and filled out an acceptability survey. All unused medications were disposed of by the study team.

Software and Hardware

CPMRx software delivers a user-friendly platform that (1) allows users to report dose-by-dose pain scores, (2) helps users consider whether a dose is needed, and (3) creates use traceability. The software allows a user to report their pain score using a modified Visual Analog Scale when a dose is taken and includes a user-directed "gamification" component. This component delivers positive reinforcement cues to the user for managing their pain within recommended treatment protocols with the goal of providing education and incentivizing patients to make smarter and more informed decisions about dose frequency and amount. The software collects and organizes data that can be accessed by clinicians to view trending analysis for pain scores, adherence to treatment plans, and time between doses.

For this study, the mobile app was installed on study-provided smart tablets that were locked into "kiosk mode" (ie, altered to only provide access to the CPMRx mobile app and necessary system functions). The EM pill blister packs used in this study were manufactured by Information Mediary Corporation. Compliance data were extracted from blister packs using a desktop radio frequency identification reader and were then transferred into an electronic record for research purposes.

Measures

Sociodemographic variables were extracted from patients' electronic medical records. These included age (in years), sex (male or female), race (White or European American, Black or African American, Asian American, American Indian or Alaska Native, Native Hawaiian or Pacific Islander, 2 or more races, or unknown), current smoking status (smoker or nonsmoker), and history of opioid prescription (dichotomized as yes or no).

Patients self-reported pain within the mobile app on a modified Visual Analog Scale with tick marks at 0-10. The screen reads "How much pain are you in?" and there is a face that changes colors as the user chooses the pain rating on the slider. For those with dexterity issues, the tick marks can simply be tapped.

XSL•FO
RenderX

Acceptability assessments included 2 quantitative measures and 7 open-ended questions. Quantitative measures were ease of use (rated on a 5-point scale from 1="very hard" to 5="very easy") and amount of time required (rated on a 5-point scale from 1="too much" to 5="very little"). Open-ended questions sought to better understand the overall user experience and possibilities for improvement (eg, "What did you like most about your experience with the app?" and "How would you improve this app?"). Demand and engagement were measured as the actual use of the mobile app during the postoperative period, and logistical and technological feasibility were assessed by reports of issues with app use. The total number of opioids used was also measured to determine their potential use in informing clinical decision-making.

Data Analysis

All quantitative analyses were run using SAS (version 9.4; SAS Institute). Univariate statistics were used to summarize both continuous (eg, means, SDs, and medians) and categorical (eg, proportions and total numbers) variables for sample characteristics. Descriptive statistics were run to determine means and SDs of quantitative measures of acceptability.

Qualitative data were coded through thematic analysis using NVivo (version 14; Lumivero). Due to the simplicity of the text, a single researcher coded all open-ended responses and generated themes using an inductive semantic approach.

Ethical Considerations

This study was approved by the institutional review board at the CGVAMC and was conducted in the performance of a cooperative research and development agreement between Continuous Precision Medicine and the Veterans Health Administration Innovation Ecosystem. All participants provided informed consent and HIPAA authorization. All study data were deidentified. Participants were provided a US $50 gift card for their participation.

Results

Sample Characteristics

The sample for this pilot was quite homogeneous; 9 (90%) of the veterans were male, and all 10 (100%) were White nonsmokers with no history of prior opioid prescription. The mean age of the participants was 68.8 (SD 9.7) years with a range of 54 to 81 years.

Feasibility Outcomes

Overall, quantitative measures of acceptability were high. The average rating for the amount of time required to use the app was 4.88 (SD 0.35; range 4.0-5.0), and the average rating for ease of use was 4.38 (SD 1.06). One participant (age 77 years, male) rated ease of use of the mobile app as hard (2 of 5). Thematic analysis of open-ended questions also revealed ease of use as a central theme, and that most participants found ease of use to be high. For instance, in response to the question, "What did you like most about your experience with the app?" participants cited "ease of documentation," "very straightforward," "easy to use," "easy to understand," and "increases awareness of medication use and very easy to use." Demand and engagement were high as well, with a mean number of mobile app entries of 34.1 (SD 20.1) during the postoperative period. There were no reported technological or logistical issues with the mobile app. Participants took an average of 25.13 (SD 14.37) opioid tablets to manage their postoperative pain.

Discussion

The results of this study revealed that the use of a mobile app for pain and medication management during postoperative recovery was both feasible and acceptable in older veterans undergoing total knee arthroplasty within the Veterans Health Administration. Overall, participants rated ease of use and amount of time required as highly acceptable, with no logistical or technological issues being reported. This is congruent with our prior findings across several studies, including those involving young soldiers undergoing third molar extraction [24,25], adults undergoing tonsillectomy [26], and women undergoing cesarean section [27,28]. This study was the first to include a report by a patient of difficulty using the mobile app, which does warrant discussion. As mentioned previously, a male participant aged 77 years reported that he found the mobile app hard to use and noted in his comments that it was "not intuitive." Although there is a tutorial embedded within the app, it may be that an additional point of socialization to the software in this population may be helpful. It is our recommendation that the clinical point of contact (eg, the orthopedic nurse case manager) demonstrate how the patient should record information within the app prior to discharge, particularly if the patient reports unfamiliarity with mobile app technologies. This in-person tutorial should mitigate any acceptability concerns for the patient and would, in our estimation, require less than 2 additional minutes of the nurse's time.

Demand for and engagement with the mobile app were high, which supports prior research among veterans suggesting an interest in adopting mHealth technologies [16]. The number of entries by participants showed that repeated inputs (at each time of medication use) were not too burdensome over the course of the 7-day prescription. In fact, more entries were logged over this 7-day postoperative period than in prior pain studies [22]. Finally, the wide variation in opioid consumption across participants revealed the potential use of the mobile app to provide actionable insights to clinicians if adopted more widely. Taken together, these results suggest that scaling use of the CPMRx mobile app in this population is both feasible and acceptable. Given the small sample size, generalizability of findings is limited. Future research should examine whether these patterns hold across surgical services within the Veterans Health Administration and for longer lengths of care as would be the case for patients dealing with chronic pain. Implementing the CPMRx suite of software may improve clinical care by reducing opioid misuse and preventing habit-forming behaviors at the individual level as well as decreasing excess prescription opioids available for diversion and thereby reducing risk at the community level.

XSL•FO
RenderX

Acknowledgments

This material is based upon work supported in part by the North Carolina Biotechnology Center and conducted in the performance of a cooperative research and development agreement between Continuous Precision Medicine, the Veterans Health Administration Innovation Ecosystem, and the Western North Carolina Veterans Affairs Health Care System. This material is the result of work supported with resources and the use of facilities at the Charles George VA Medical Center, Asheville, North Carolina. We would like to thank Matthew D Hollifield, PharmD, BCPS, clinical pharmacist, Charles George VA Medical Center, for his assistance in data collection preparation and logistics.

Data Availability

The data sets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Disclaimer

The contents of this publication do not represent the views of the US Department of Veterans Affairs or the US government.

Conflicts of Interest

JKM, SKW, AH, and AO are employees of Continuous Precision Medicine. To mitigate the potential for perceived conflict, all data collection was conducted by members of the Charles George VA Medical Center study team.

References

1. Key substance use and mental health indicators in the United States: results from the 2019 National Survey on Drug Use and Health (HHS publication no. PEP20-07-01-001, NSDUH series H-55). Substance Abuse and Mental Health Services Administration. 2020. URL: https://store.samhsa.gov/product/key-substance-use-and-mental-health-indicators-in-the-united-states-results-from-the-2019-national-survey-on-Drug-Use-and-Health/PEP20-07-01-001 [accessed 2023-09-28]
2. Ahmad F, Rossen L, Sutton P. Provisional drug overdose death counts. US Centers for Disease Control and Prevention. 2021. URL: https://www.cdc.gov/nchs/nvss/vsrr/drug-overdose-data.htm [accessed 2023-09-28]
3. Overdose deaths accelerating during COVID-19. US Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/media/releases/2020/p1218-overdose-deaths-covid-19.html [accessed 2023-09-28]
4. Kelley ML, Bravo AJ, Votaw VR, Stein E, Redman JC, Witkiewitz K. Opioid and sedative misuse among veterans wounded in combat. Addict Behav 2019;92:168-172 [FREE Full text] [doi: 10.1016/j.addbeh.2018.12.007] [Medline: 30640149]
5. Florence CS, Zhou C, Luo F, Xu L. The economic burden of prescription opioid overdose, abuse, and dependence in the United States, 2013. Med Care 2016;54(10):901-906 [FREE Full text] [doi: 10.1097/MLR.0000000000000625] [Medline: 27623005]
6. Florence C, Luo F, Rice K. The economic burden of opioid use disorder and fatal opioid overdose in the United States, 2017. Drug Alcohol Depend 2021;218:108350 [FREE Full text] [doi: 10.1016/j.drugalcdep.2020.108350] [Medline: 33121867]
7. Baser O, Xie L, Mardekian J, Schaaf D, Wang L, Joshi AV. Prevalence of diagnosed opioid abuse and its economic burden in the Veterans Health Administration. Pain Pract 2014;14(5):437-445 [FREE Full text] [doi: 10.1111/papr.12097] [Medline: 23809020]
8. Brummett CM, Waljee JF, Goesling J, Moser S, Lin P, Englesbe MJ, et al. New persistent opioid use after minor and major surgical procedures in US adults. JAMA Surg 2017;152(6):e170504 [FREE Full text] [doi: 10.1001/jamasurg.2017.0504] [Medline: 28403427]
9. Compton WM, Jones CM, Baldwin GT. Relationship between nonmedical prescription-opioid use and heroin use. N Engl J Med 2016;374(2):154-163 [FREE Full text] [doi: 10.1056/NEJMra1508490] [Medline: 26760086]
10. Banerjee G, Edelman EJ, Barry DT, Becker WC, Cerdá M, Crystal S, et al. Non-medical use of prescription opioids is associated with heroin initiation among US veterans: a prospective cohort study. Addiction 2016;111(11):2021-2031 [FREE Full text] [doi: 10.1111/add.13491] [Medline: 27552496]
11. Han B, Compton WM, Blanco C, Crane E, Lee J, Jones CM. Prescription opioid use, misuse, and use disorders in U.S. adults: 2015 National Survey on Drug Use and Health. Ann Intern Med 2017;167(5):293-301 [FREE Full text] [doi: 10.7326/M17-0865] [Medline: 28761945]
12. Hill MV, McMahon ML, Stucke RS, Barth RJ. Wide variation and excessive dosage of opioid prescriptions for common general surgical procedures. Ann Surg 2017;265(4):709-714 [FREE Full text] [doi: 10.1097/SLA.0000000000001993] [Medline: 27631771]

XSL•FO
RenderX

13. Hill MV, Stucke RS, Billmeier SE, Kelly JL, Barth RJ. Guideline for discharge opioid prescriptions after inpatient general surgical procedures. J Am Coll Surg 2018;226(6):996-1003. [doi: 10.1016/j.jamcollsurg.2017.10.012] [Medline: 29198638]
14. Hill MV, Stucke RS, McMahon ML, Beeman JL, Barth RJ. An educational intervention decreases opioid prescribing after general surgical operations. Ann Surg 2018;267(3):468-472. [doi: 10.1097/SLA.0000000000002198] [Medline: 28267689]
15. Demographics of mobile device ownership and adoption in the United States. Pew Research Center. 2021. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2022-01-27]
16. Erbes CR, Stinson R, Kuhn E, Polusny M, Urban J, Hoffman J, et al. Access, utilization, and interest in mHealth applications among veterans receiving outpatient care for PTSD. Mil Med 2014;179(11):1218-1222 [FREE Full text] [doi: 10.7205/MILMED-D-14-00014] [Medline: 25373044]
17. Tofighi B, Abrantes A, Stein MD. The role of technology-based interventions for substance use disorders in primary care: a review of the literature. Med Clin North Am 2018;102(4):715-731 [FREE Full text] [doi: 10.1016/j.mcna.2018.02.011] [Medline: 29933825]
18. Scott CK, Dennis ML, Gustafson DH. Using smartphones to decrease substance use via self-monitoring and recovery support: study protocol for a randomized control trial. Trials 2017;18(1):374 [FREE Full text] [doi: 10.1186/s13063-017-2096-z] [Medline: 28797307]
19. Ploug T, Holm S. Eliciting meta consent for future secondary research use of health data using a smartphone application—a proof of concept study in the Danish population. BMC Med Ethics 2017;18(1):51 [FREE Full text] [doi: 10.1186/s12910-017-0209-6] [Medline: 28810914]
20. Frank E, Pong J, Asher Y, Soares CN. Smart phone technologies and ecological momentary data: is this the way forward on depression management and research? Curr Opin Psychiatry 2018;31(1):3-6. [doi: 10.1097/YCO.0000000000000382] [Medline: 29084010]
21. Santo K, Chow CK, Thiagalingam A, Rogers K, Chalmers J, Redfern J. MEDication reminder APPs to improve medication adherence in Coronary Heart Disease (MedApp-CHD) study: a randomised controlled trial protocol. BMJ Open 2017;7(10):e017540 [FREE Full text] [doi: 10.1136/bmjopen-2017-017540] [Medline: 28993388]
22. Jamison RN, Jurcik DC, Edwards RR, Huang CC, Ross EL. A pilot comparison of a smartphone app with or without 2-way messaging among chronic pain patients: who benefits from a pain app? Clin J Pain 2017;33(8):676-686 [FREE Full text] [doi: 10.1097/AJP.0000000000000455] [Medline: 27898460]
23. Park LG, Howie-Esquivel J, Dracup K. Electronic measurement of medication adherence. West J Nurs Res 2015;37(1):28-49. [doi: 10.1177/0193945914524492] [Medline: 24577868]
24. Rethman M, Morgan JK, Moorhead-Beardsley B, Walther SK, Arnold JP. Opioid use and misuse among active-duty soldiers following postoperative prescription: using predictors to move toward precision prescribing practices. 2022 Presented at: Military Health System Research Symposium; September 5, 2022; Kissimmee, FL URL: https://www.researchgate.net/publication/362407942_Predictors_of_Opioid_Use_and_Misuse_among_Active-Duty_Soldiers_Following_Postoperative_Prescription#:~:text=4%20and%20later.-,Overall%2C%20the%20model%20predicting%20total%20opioid%20use%20was%20significant%20and,age%20emerged%20as%20significant%20predictors
25. Morgan JK, Rethman M, Moorhead-Beardsley B, Walther SK, Arnold JP. Predictors of opioid use and misuse among active-duty soldiers following postoperative prescription. 2022 Presented at: American Psychological Association Annual Convention; August 2022; Minneapolis, MN.
26. Morgan JK, Walther SK, Lane ME. Evaluation of a mobile application to decrease opioid misuse and habit-forming behaviors following prescription: preliminary results and future directions. Psychology 2019;10(15):2019-2025 [FREE Full text] [doi: 10.4236/psych.2019.1015129]
27. Gutta S, Morgan JK, Hart L, Abdo A, Leib A, Walther SK, et al. Feasibility and acceptability of a mobile application to monitor opioid use and promote behavior modification following cesarean section. Psychology 2022;13(3):375-387 [FREE Full text] [doi: 10.4236/psych.2022.133024]
28. Morgan J, Abdo A, O'Connor M, Hart L, Leib A, Walther S, et al. Evaluation of a mobile application to decrease opioid misuse in patients undergoing cesarean section. 2023 Presented at: American Psychological Association Annual Convention; August 2, 2023; Washington, DC URL: https://www.researchgate.net/publication/372853636_Evaluation_of_a_Mobile_Application_to_Decrease_Opioid_Misuse_in_Patients_Undergoing_Cesarean_Section

Abbreviations

CGVAMC: Charles George VA Medical Center
EM: electronic monitoring
HIPAA: Health Insurance Portability and Accountability Act
mHealth: mobile health

XSL•FO
RenderX

Edited by T de Azevedo Cardoso; submitted 19.06.23; peer-reviewed by D Dunsmuir; comments to author 18.07.23; revised version received 11.08.23; accepted 21.09.23; published 18.10.23.

Please cite as:
Morgan JK, Rawlins CR, Walther SK, Harvey A, O'Donnell A, Greene M, Schmidt TG
A Mobile App for Postoperative Pain Management Among Older Veterans Undergoing Total Knee Arthroplasty: Mixed Methods Feasibility and Acceptability Pilot Study
JMIR Perioper Med 2023;6:e50116
URL: https://periop.jmir.org/2023/1/e50116
doi:10.2196/50116
PMID:37851497

©Jessica Kelley Morgan, Caitlin R Rawlins, Steven K Walther, Andrew Harvey, Annmarie O'Donnell, Marla Greene, Troy G Schmidt. Originally published in JMIR Perioperative Medicine (http://periop.jmir.org), 18.10.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Perioperative Medicine, is properly cited. The complete bibliographic information, a link to the original publication on http://periop.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Description of the Content and Quality of Publicly Available Information on the Internet About Inhaled Volatile Anesthesia and Total Intravenous Anesthesia: Descriptive Study

Xinwen Hu[1], MD, MPHS; Bethany R Tellor Pennington[1], PharmD, BCPS; Michael S Avidan[1], MBBCh; Sachin Kheterpal[2], MD, MBA; Nastassjia G deBourbon[3], MD; Mary C Politi[4], PhD

[1]Department of Anesthesiology, Washington University in St Louis School of Medicine, St Louis, MO, United States

[2]Department of Anesthesiology, University of Michigan, Ann Arbor, MI, United States

[3]Magnolia Regional Health Center, Corinth, MS, United States

[4]Department of Surgery, Washington University in St Louis School of Medicine, St Louis, MO, United States

Corresponding Author:
Xinwen Hu, MD, MPHS
Department of Anesthesiology
Washington University in St Louis School of Medicine
660 S Euclid Avenue
MSC 8054-0043-12
St Louis, MO, 63110
United States
Phone: 1 3143628649
Email: huxinwen@wustl.edu

Abstract

Background: More than 300 million patients undergo surgical procedures requiring anesthesia worldwide annually. There are 2 standard-of-care general anesthesia administration options: inhaled volatile anesthesia (INVA) and total intravenous anesthesia (TIVA). There is limited evidence comparing these methods and their impact on patient experiences and outcomes. Patients often seek this information from sources such as the internet. However, the majority of websites on anesthesia-related topics are not comprehensive, updated, and fully accurate. The quality and availability of web-based patient information about INVA and TIVA have not been sufficiently examined.

Objective: This study aimed to (1) assess information on the internet about INVA and TIVA for availability, readability, accuracy, and quality and (2) identify high-quality websites that can be recommended to patients to assist in their anesthesia information-seeking and decision-making.

Methods: Web-based searches were conducted using Google from April 2022 to November 2022. Websites were coded using a coding instrument developed based on the International Patient Decision Aids Standards criteria and adapted to be appropriate for assessing websites describing INVA and TIVA. Readability was calculated with the Flesch-Kincaid (F-K) grade level and the simple measure of Gobbledygook (SMOG) readability formula.

Results: A total of 67 websites containing 201 individual web pages were included for coding and analysis. Most of the websites provided a basic definition of general anesthesia (unconsciousness, n=57, 85%; analgesia, n=47, 70%). Around half of the websites described common side effects of general anesthesia, while fewer described the rare but serious adverse events, such as intraoperative awareness (n=31, 46%), allergic reactions or anaphylaxis (n=29, 43%), and malignant hyperthermia (n=18, 27%). Of the 67 websites, the median F-K grade level was 11.3 (IQR 9.5-12.8) and the median SMOG score was 13.5 (IQR 12.2-14.4), both far above the American Medical Association (AMA) recommended reading level of sixth grade. A total of 51 (76%) websites distinguished INVA versus TIVA as general anesthesia options. A total of 12 of the 51 (24%) websites explicitly stated that there is a decision to be considered about receiving INVA versus TIVA for general anesthesia. Only 10 (20%) websites made any direct comparisons between INVA and TIVA, discussing their positive and negative features. A total of 12 (24%) websites addressed the concept of shared decision-making in planning anesthesia care, but none specifically asked patients to think about which features of INVA and TIVA matter the most to them.

XSL•FO
RenderX

Conclusions: While the majority of websites described INVA and TIVA, few provided comparisons. There is a need for high-quality patient education and decision support about the choice of INVA versus TIVA to provide accurate and more comprehensive information in a format conducive to patient understanding.

(JMIR Perioper Med 2023;6:e47714) doi:10.2196/47714

KEYWORDS

internet; general anesthesia; inhaled volatile anesthesia; total intravenous anesthesia; patient education; shared decision-making; surgery; information; decision-making; web-based; anesthesia; anesthesiology; anesthesiologist

Introduction

More than 300 million patients undergo surgical procedures requiring anesthesia worldwide annually [1]. Inhaled volatile anesthesia (INVA) and total intravenous anesthesia (TIVA) are the 2 most commonly used standard-of-care general anesthesia administration methods. Insufficient evidence is available to establish which method is associated with superior patient experiences and outcomes. In the absence of robust comparative effectiveness trials evaluating patient experiences with each option, it is likely that most clinicians feel unable to discuss the differences from the patient's perspective between these general anesthesia techniques, leaving patients who are interested in this comparison to seek information from the internet or other sources. For example, a recent survey noted that 40% of patients who have had surgery in the past 5 years were not included in the decision to choose INVA versus TIVA and almost half of these patients looked for information on their own about general anesthesia before their surgery. Of the 585 places searched, 412 (70%) were online websites [2]. Many patients report using the internet to learn more about their surgical procedures in general [3-5]. Enabling patients to be informed with the best available evidence is a critical component of high-quality patient care [6]. Information gathered from web-based sources can influence patients' decision-making [7], so it is important to ensure patients are able to access accurate, comprehensible, and high-quality information.

Unfortunately, the majority of the websites on anesthesia-related topics are not comprehensive, updated, and fully accurate [8-15]. In addition, although some high-quality and accurate websites about anesthesia exist, they do not always rise to the top of the search engine results [14]. Moreover, most websites on anesthesia-related topics are written at a reading level above the American Medical Association (AMA) recommended level of sixth grade [8,10,11,15,16]. In fact, the median readability level is around a 13.5 (IQR 12.0-14.6) grade level, at which only about 62% of US adults can easily understand [11,17]. There are some available website quality certifications developed by national organizations or independent foundations, but these are not always indicative of content quality [10]. In addition, patients have expressed concerns with the process of searching for web-based information about general anesthesia. In a previous survey, 65% of patients who sought information about general anesthesia through web-based resources noted that it took a lot of effort to get the information they needed; 53% felt frustrated during their search [2].

Current studies have not sufficiently examined the quality and availability of web-based patient information about INVA and TIVA. The objective of this study was to assess publicly available information on the internet regarding both methods of general anesthesia administration for availability, readability, accuracy, and quality. We aimed to identify high-quality websites that can be recommended to patients to help assist them in making informed decisions about their anesthetic care.

Methods

Website Selection

Web-based searches were conducted using Google, the most commonly used search engine worldwide with a market share of over 90% [18-20]. All searches were performed in the United States from April 2022 to November 2022, using a browser with no stored cookies or browsing history to avoid generating personalized results. The following keywords were searched: "general anaesthesia," "general anesthesia," "anaesthesia," "anesthesia," "putting to sleep for surgery," "propofol," "intravenous anaesthesia," "intravenous anesthesia," "total intravenous anaesthesia," "total intravenous anesthesia," "inhaled volatile anaesthesia," "inhaled volatile anesthesia," "anaesthetic gases," and "anesthetic gases."

Each web page had to meet all of the following eligibility criteria to be included: (1) was displayed within the first 3 pages of search engine results when searching any of the keywords specified above, as over 90% of individuals do not look beyond the first 3 pages [19,20]; (2) was publicly available with no login required; (3) contained information on general anesthesia; (4) was intended for adult surgical patients; and (5) was written in English. Web pages were excluded if any of the following criteria existed: (1) they required logins, including subscriptions or free sign-ups; (2) they did not contain information about general anesthesia; (3) they were targeted toward medical professionals, defined as either websites that explicitly stated they were intended for use by medical professionals, or search results linked to books or scholarly journal papers that were not labeled as patient information pages; (4) they were written for pediatric patients and their parents; (5) they had a primary format of the video, social media, discussion board, question and answer forum, chat room, or personal blog; (6) they were identified by Google as a sponsored advertisement; or (7) they were written in any language besides English.

Each included web page and pages with the same domain name linked within 2 clicks were considered as a single website for subsequent coding and analysis. Linked web pages were excluded from the analysis if any of the exclusion criteria existed. External sites or references linked from eligible web

XSL•FO
RenderX

pages were excluded. All embedded videos were excluded from the analysis.

Ethical Considerations

As this study did not involve human subjects' research, institutional review board oversight was not needed.

Website Coding

A coding instrument (Multimedia Appendix 1 [21-41]) was developed based on the International Patient Decision Aids Standards (IPDAS) criteria for high-quality patient decision tools [21,22], adapted to be appropriate for measuring the quality of websites describing INVA and TIVA. Part I of the coding instrument was applied to all included websites that contained information on general anesthesia. This portion contained 28 items from the following categories: (1) basic definition and description of general anesthesia, (2) side effects and potential harms of general anesthesia, (3) what to expect with general anesthesia during the perioperative period, and (4) whether the website describes both inhaled and intravenous anesthesia as general anesthesia options. Items in part I included selected items from the IPDAS minimum standards of information necessary to support decisions and additional items adapted from previous research examples [19,20,42]. Part II of the coding instrument was adapted from the remaining IPDAS quality criteria needed to improve the quality of patient materials and was developed to evaluate the subset of websites that discussed both intravenous and inhaled anesthetic options. This section contained 29 items from the following areas: (1) comparison between INVA and TIVA, (2) qualitative or quantitative description of adverse event probabilities, (3) guidance for choosing between INVA and TIVA, and (4) evidence selection and disclosure. The coding instrument was discussed and iteratively revised by the study team to ensure clarity and agreement on definitions, coding approach, and items included. Once the team agreed on the items and coding process, the coder (XH) coded a sample website and clarified the remaining questions before coding the identified websites for analysis.

An item was checked if a website presented corresponding information in an accurate way, or if the criterium were satisfied per the rater's judgment. All website coding was performed by a single researcher (XH) given the quantitative nature of the coding structure. Any ambiguity about coding was discussed among 3 of the authors (XH, BRTP, and MCP) and final decisions were made by consensus.

Readability Assessment

The URLs of all web pages of each website were submitted to ReadablePro [43] for readability score calculation. Only the main body of the text was analyzed; header, footer, and references were excluded from the analysis. Readability for each web page was calculated with the Flesch-Kincaid (F-K) grade level [44] and the simple measure of Gobbledygook (SMOG) readability formula [45]. We randomly selected 10 web pages and manually calculated the F-K and SMOG readability to check accuracy; results from the ReadablePro calculation were consistent with manual calculations.

Data Analysis

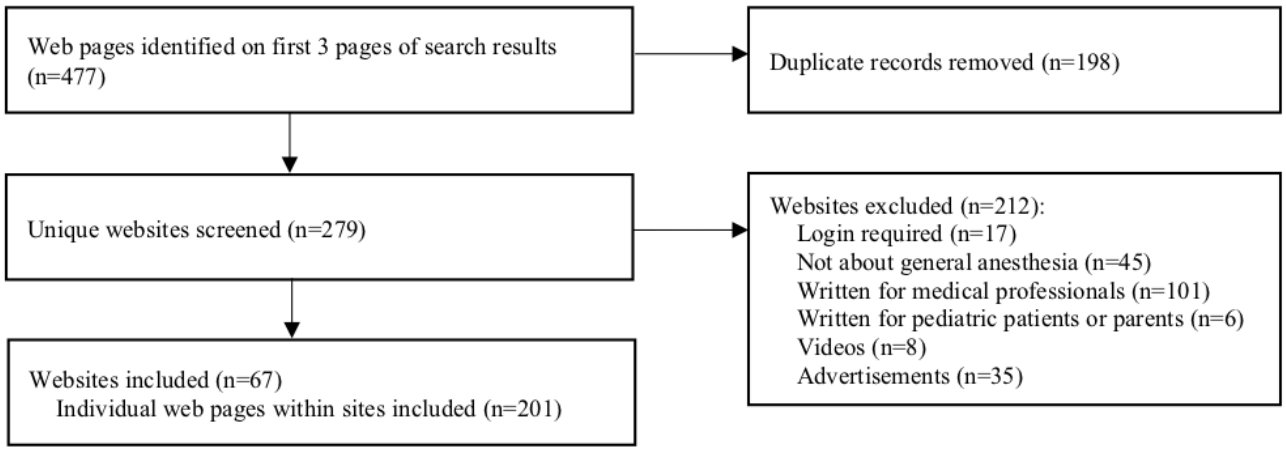
The frequency and percentage of websites that checked each item in the coding instrument were calculated. For each website, the number of items it checked in each category of part I and part II of the coding instrument were tabulated to determine the most comprehensive websites. For the readability of each website, the mean F-K grade level and mean SMOG score across all of its web pages were calculated. Descriptive statistics, including median, range, IQR, were then calculated for website mean F-K grade levels and website mean SMOG scores.

Results

Website Selection

The website selection process is illustrated in Figure 1. A total of 477 web pages were identified on the first 3 pages of the search results. Of those, 198 duplicate records were removed. Of the remaining 279 websites, 212 were excluded because they required logins (n=17), did not contain information on general anesthesia (n=45), were targeted toward medical professionals (n=101), were written for pediatric patients and their parents (n=6), had a primary format of a video (n=8), or were identified by Google as advertisements (n=35). A total of 67 websites were formed from the included web pages and eligible pages linked within 2 clicks, with a total of 201 individual web pages included for coding and analysis.

Figure 1. Inclusion or exclusion of websites identified in searches.



XSL•FO
RenderX

Website Coding

Content on General Anesthesia

Overall Description of General Anesthesia

All 67 websites were assessed using part I of the coding instrument (Table 1). Most of the websites provided a basic definition of general anesthesia (unconsciousness, n=57, 85%; analgesia, n=47, 70%) and described who administers general anesthesia (n=51, 76%). However, fewer than half of the websites discussed how general anesthesia is monitored during surgery (described monitoring of vital signs, n=33, 49%; described monitoring of "level of unconsciousness or awareness," n=21, 31%; mentioned brain monitoring specifically, n=9, 13%).

Few websites discussed how general anesthesia works. Only 21 out of the 67 (31%) websites mentioned that general anesthesia provides control of the airway and breathing and allows for surgeries that affect breathing. A total of 17 (25%) websites mentioned muscle relaxation or immobility creates a controlled operative condition, 28 (42%) websites described the fact that general anesthesia works rapidly, and only 11 (16%) described the role of general anesthesia for surgeries that take a long time requiring longer sedation.

XSL•FO
RenderX

Table 1. Content of websites about general anesthesia (N=67).

Items	Number of websites, n (%)
Basic definition and description of general anesthesia	
Provided basic definition of general anesthesia	
Unconsciousness	57 (85)
Analgesia	47 (70)
Discussed who administers general anesthesia	51 (76)
Discussed how general anesthesia is monitored during surgery	
Monitoring of vital signs	33 (49)
Assessment of "level of unconsciousness/awareness"	21 (31)
Brain monitoring (eg, processed electroencephalogram monitoring)	9 (13)
Provides control of airway and breathing or allows for surgeries that affect breathing	21 (31)
Provides muscle relaxation or immobility to prevent involuntary movements and create a controlled operative condition	17 (25)
Has a rapid onset of effect	28 (42)
Allows for surgeries that take a long time	11 (16)
Side effects and potential harms of general anesthesia	
Common side effects	
PONV[a]	41 (61)
Chills or shivering	23 (34)
Sleepiness or confusion	34 (51)
Changes in heart rate and blood pressure	30 (45)
Rare but serious adverse events	
Intraoperative awareness	31 (46)
Allergic reaction or anaphylaxis	29 (43)
Malignant hyperthermia	18 (27)
Propofol related infusion syndrome	10 (15)
Risk factors for general anesthesia adverse events	
Risk factors for PONV	9 (13)
Risk factors for intraoperative awareness	17 (25)
Risk factors for malignant hyperthermia	9 (13)
What to expect before, during, and after surgery with general anesthesia	
Before surgery, patient will meet with their anesthesia care team to…	
Review medical history	30 (45)
Discuss anesthesia options	23 (34)
Patient will need to fast before surgery	27 (40)
After anesthesia is administered, the patient will receive an endotracheal tube or alternative airway options	37 (55)
After the surgery is completed, anesthesia will be discontinued, and patient will regain consciousness	22 (33)
Some patients may take a longer time to wake up	2 (3)
Patient may have worse pain as the anesthesia wears off	9 (13)
Description of inhaled and intravenous anesthetics as general anesthesia options	
Described inhaled and intravenous anesthetics as general anesthesia options	51 (76)

[a]PONV: postoperative nausea and vomiting.

XSL•FO
RenderX

Side Effects and Potential Harms of General Anesthesia

Around half of the websites described common side effects of general anesthesia such as postoperative nausea and vomiting (PONV; n=41, 61%), chills or shivering (n=23, 34%), sleepiness or confusion (n=34, 51%), and changes in heart rate and blood pressure (n=30, 45%). For rare but serious adverse events of general anesthesia, intraoperative awareness was described by 31 (46%) websites, allergic reactions or anaphylaxis by 29 (43%), malignant hyperthermia by 18 (27%), and propofol related infusion syndrome by 10 (15%). At least 1 risk factor for PONV was mentioned by 9 (13%) websites, 17 (25%) discussed risk factors for intraoperative awareness, and 9 (13%) discussed those for malignant hyperthermia.

Expectations for the Perioperative Period

Fewer than half of the websites described what to expect before surgery, including meeting with their anesthesia team to review medical history (n=30, 45%) and discuss anesthesia options (n=23, 34%) and fasting before surgery (n=27, 40%). The probable need for endotracheal intubation or alternative airway options was discussed by 37 (55%) websites. A total of 22 (33%) websites explicitly stated that anesthesia will be discontinued at the end of surgery for patients to regain consciousness, but only 2 (3%) mentioned the possibility of needing a longer time to regain consciousness. Only 9 (13%) websites helped set the expectation that patients may experience worsening pain as anesthesia wears off.

Content on INVA versus TIVA

Overview

Of the 67 websites analyzed, 51 (76%) distinguished inhaled versus intravenous anesthetics as general anesthesia options. These websites were further assessed with part II of the coding instrument (Table 2).

XSL•FO
RenderX

Table 2. Content of websites that described inhaled volatile anesthesia versus total intravenous anesthesia as general anesthetic options (N=51).

Items	Number of websites, n (%)
Information	
Explicitly stated there is a decision that needs to be considered regarding using INVA[a] or TIVA[b] when general anesthesia is indicated	12 (24)
Mentioned the decision of using INVA vs TIVA depends on…	
Clinician's preference	0 (0)
Patient's medical history	4 (8)
Surgery or procedure requirements	1 (2)
Patient's preferences	0 (0)
Described positive features of INVA	
Standard of care for decades	8 (16)
Predictable dose-response relationship	3 (6)
Described positive features of TIVA	
Standard of care for decades	7 (14)
Lower risk of postoperative nausea and vomiting compared to inhaled agents	8 (16)
Described negative features of INVA	
Malignant hyperthermia	9 (18)
Greenhouse gases or more atmospheric pollution compared to TIVA	6 (12)
Described negative features of TIVA	
Reactions to propofol, for example, allergic reaction, anaphylaxis, bacterial contamination leading to infection, and propofol infusion syndrome	9 (18)
Compared the costs of INVA and TIVA	4 (8)
Showed the negative and positive features of the 2 general anesthesia administration options with equal detail	12 (24)
Made it possible to compare the positive and negative features of INVA versus TIVA	10 (20)
Probabilities	
Provided numeric or qualitative descriptions of the probabilities of the adverse effects associated with INVA and TIVA	1 (2)
Provided more than 1 way of viewing the probabilities	0 (0)
Provided information about the levels of uncertainty around adverse event probabilities	1 (2)
Values	
Explicitly stated shared decision making is an option for anesthesia	12 (24)
Asked patients to think about what matters most to them	0 (0)
Guidance	
Provided a step-by-step way to choose anesthesia method	0 (0)
Included tools like worksheets or list of questions to use when discussing anesthesia options with a clinician	5 (10)
Evidence	
Provided citations to the evidence selected	24 (47)
Provided the date of publication or the date of last update	35 (69)
Provided information about the update policy	3 (6)
Disclosure	
Stated funding source and institutional affiliations	51 (100)
Provided author or medical reviewer credentials	20 (39)

XSL•FO
RenderX

[a]INVA: inhaled volatile anesthesia.

[b]TIVA: total intravenous anesthesia.

Information Criteria

A total of 12 out of the 51 (24%) websites explicitly stated that there is a decision to be considered regarding whether to use INVA or TIVA when general anesthesia is indicated. A minority of these websites explained that such a decision depends on patients' medical history (n=4, 8%) or surgery or procedure requirements (n=1, 2%), and no website mentioned that this decision may also factor in clinicians' or patients' preferences.

INVA was identified as standard of care by 8 (16%) websites and TIVA by 7 (14%). In terms of the pros and cons of INVA and TIVA, only 10 out of the 51 (20%) websites made any direct comparisons between the 2 general anesthesia options. A total of 8 (16%) websites mentioned TIVA is associated with a lower risk of PONV compared to INVA. A total of 3 (6%) websites explored the more predictable dose-response relationship of INVA. Few websites specifically associated malignant hyperthermia (n=9, 18%) or worse atmospheric pollution (n=6, 12%) with INVA, and only 9 (18%) mentioned any adverse drug reactions. A total of 4 (8%) websites provided information related to the costs of each method.

A total of 28 of the 51 (55%) websites did not provide any further information about INVA and TIVA beyond distinguishing between inhaled and intravenous anesthetics. A total of 11 (22%) websites provided some information about each option, but failed to do so in a balanced way with similar amounts of information for each. Only 12 (24%) websites discussed INVA and TIVA with enough detail to distinguish between them, and presented information about the 2 anesthetic options with an equal amount of detail.

Probabilities Criteria

Only 1 website provided a qualitative description of the probabilities of adverse effects associated with INVA or TIVA. It expressed the level of uncertainty around the adverse event probabilities, but failed to provide alternative ways of viewing the probabilities (such as graphs). None of the websites provided any quantitative description of adverse event probabilities associated with INVA or TIVA.

Values and Guidance Criteria

A total of 12 out of the 51 (24%) websites mentioned the concept of shared decision-making in planning anesthesia care, but none asked patients to think about which features of INVA and TIVA matter the most to them in the specific setting of choosing between the 2. No websites provided any step-by-step way to guide patients in choosing which general anesthesia administration method they prefer. Of the 51 websites, 5 (10%) websites provided a list of questions that patients can ask when discussing their anesthesia care with clinicians, but none of those questions was specifically developed to facilitate the discussion with clinicians about choosing between INVA and TIVA.

Evidence and Disclosure Criteria

A total of 24 (47%) websites provided citations to the evidence selected. A total of 35 (69%) provided the date of publication or the date of the last update, but only 3 (6%) provided information about their update policy to help patients assess whether the information is outdated. All websites (n=51, 100%) disclosed their institution affiliations and funding source and 20 (39%) provided the credentials of the authors or the medical reviewers.

Readability Assessment

Website F-K grade level and SMOG score are summarized in Table 3. Of the 67 websites, the median F-K grade level was 11.3 (IQR 9.5-12.8; range 6.5-17.3), and the median SMOG score was 13.5 (IQR 12.2-14.4; range 10.3-19.0). All websites had readability levels above the AMA recommended level of sixth grade [16]. A considerable portion of websites (21% per F-K grade level and 57% per SMOG score) had readability levels ≥13, at which ≥38% of the adult population in the United States would have difficulty reading [17].

Table 3. Readability level of websites.

Readability	All websites (n=67)		Websites that distinguished inhaled from intravenous anesthesia (n=51)	
	Website F-K[a] grade level	Website SMOG[b] score	Website F-K grade level	Website SMOG score
Median readability level, median (IQR); (range)	11.3 (9.5-12.8); (6.5-17.3)	13.5 (12.2-14.4); (10.3-19.0)	11.2 (9.5-12.7); (6.5-15.6)	13.3 (12.0-14.3); (10.5-17.0)
Number (%) of websites with readability level ≥13, n (%)	14 (21)	38 (57)	11 (22)	28 (55)

[a]F-K: Flesch-Kincaid.

[b]SMOG: simple measure of Gobbledygook.

Most Comprehensive Websites

Websites that checked the highest number of items in part I of the coding instrument about general anesthesia were Wikipedia.org [46,47], verywellhealth.com [48], healthline.com [49], and GoodRx.com [50] (Table 4). For these websites, 5-17 linked web pages were included per website. Websites that were less comprehensive but had ≤3 linked web pages were ClevelandClinic.org [51,52], MedicalNewsToday.com [53], NHS.uk [54,55], NHSinform.scot [56], and OUH.NHS.uk [57]. Regarding specific subcategories of information about general anesthesia, Wikipedia.org [46,47] and verywellhealth.com [48]

XSL•FO
RenderX

covered the highest number of items in the category of a basic definition and description of general anesthesia. Websites that covered the highest number of items in the category of side effects and potential harms of general anesthesia were Wikipedia.org [46,47], GoodRx.com [50], ASAHQ.org [58], ClevelandClinic.org [51,52], Drugs.com [59,60], and verywellhealth.com [48]. In terms of the expectations for the perioperative period, websites that covered the highest number of items were verywellhealth.com [48], MayoClinic.org [61], NHS.uk [54,55], OUH.NHS.uk [57], and Patient.info [62]. Among the 51 websites that distinguished inhaled versus intravenous anesthetics, Wikipedia.org [46,47] and NYSORA.com [63] checked the highest number of information items in part II of the coding instrument about INVA versus TIVA.

Table 4. List of the most comprehensive websites.

Website name	Compared INVA[a] versus TIVA[b]	Number of web-pages	Mean F-K[c] grade level	Mean SMOG[d] score
Most comprehensive websites about general anesthesia				
Wikipedia.org [46,47]	Yes	17	12.7	14.0
Verywellhealth.com [48]	Yes	12	10.1	12.8
Healthline.com [49]	Yes	5	9.5	12.3
GoodRx.com [50]	Yes	9	9.3	11.9
Less comprehensive websites about general anesthesia but easily browsed with fewer clicks				
ClevelandClinic.org [51,52]	Yes	3	10.5	13.0
MedicalNewsToday.com [53]	Yes	3	9.6	12.0
NHS.uk [54,55]	Yes	2	9.4	12.6
NHSinform.scot [56]	Yes	1	9.1	12.5
OUH.NHS.uk [57]	Yes	1	8.9	12.2
Most comprehensive websites about a basic definition and description of general anesthesia				
Wikipedia.org [46,47]	Yes	17	12.7	14.0
Verywellhealth.com [48]	Yes	12	10.1	12.8
Most comprehensive websites about sides effects and potential harms of general anesthesia				
Wikipedia.org [46,47]	Yes	17	12.7	14.0
GoodRx.com [50]	Yes	9	9.3	11.9
ASAHQ.org [58]	Yes	13	11.3	14.1
ClevelandClinic.org [51,52]	Yes	3	10.5	13.0
Drugs.com [59]	Yes	20	9.2	10.9
Verywellhealth.com [48]	Yes	12	10.1	12.8
Most comprehensive websites about expectations for the perioperative period				
Verywellhealth.com [48]	Yes	12	10.1	12.8
MayoClinic.org [61]	Yes	1	9.4	12.0
NHS.uk [54,55]	Yes	2	9.4	12.6
OUH.NHS.uk [57]	Yes	1	8.9	12.2
Patient.info [62]	Yes	4	8.3	11.4
Most comprehensive websites for the comparison between INVA and TIVA				
Wikipedia.org [46,47]	Yes	17	12.7	14.0
NYSORA.com [63]	Yes	2	15.0	16.1

[a]INVA: inhaled volatile anesthesia.

[b]TIVA: total intravenous anesthesia.

[c]F-K: Flesch-Kincaid.

[d]SMOG: simple measure of Gobbledygook.

XSL•FO
RenderX

Discussion

Principal Results

Up to 80% of US adults have used the internet to search for health information [64], and many surgical patients have looked for information on general anesthesia on their own using web-based information [2]. Consistent with previous studies on websites on anesthesia-related topics [8-15], this study identified limitations in the availability of web-based information about general anesthesia. Although majority of the included websites provided a basic definition and description of general anesthesia, most failed to explain the benefits and drawbacks of general anesthesia compared with other potentially relevant choices (eg, sedation, regional anesthesia, and local anesthesia). More importantly, most websites failed to adequately inform patients of side effects associated with general anesthesia, especially the potential for rare but serious complications. Websites often failed to describe the process of preparing for and undergoing general anesthesia during the perioperative period. The lack of information in these areas compromises the ability of web-based resources to adequately inform patients about their anesthesia care.

In addition, websites were inadequate in aiding patients in making an informed decision about receiving INVA versus TIVA for general anesthesia. INVA and TIVA are both standard-of-care general anesthesia administration methods with insufficient evidence to establish the superiority of 1 over the other regarding patient experiences and outcomes. Given this uncertainty, the choice of INVA versus TIVA is well suited to shared decision-making [65,66]. Essential elements of shared decision-making include acknowledging that there is a decision to be made, discussing the risks and benefits of available options based on best-available evidence, and eliciting patient's values and preferences [67-70]. A website useful for helping patients make an informed decision about receiving INVA versus TIVA should cover these elements. Although most websites distinguished inhaled versus intravenous anesthetics, fewer than half of them provided any further information and only a quarter explicitly stated that there is a decision to be considered regarding whether to use INVA or TIVA when general anesthesia is indicated. Although the relative advantages and disadvantages of INVA and TIVA require comparative effectiveness trials, some reliable evidence is available, which is summarized in part II of the coding instrument (Multimedia Appendix 1 [21-41]), and can be addressed by websites. However, the known comparative effectiveness evidence in relation to INVA and TIVA was presented by fewer than 20% of the websites. Only 24% of the websites described shared decision-making or incorporating patients' preferences when planning anesthesia care, and no websites asked patients to think about what is important to them when choosing between INVA versus TIVA. No websites provided additional tools, resources, or links to facilitate this decision-making process.

Of note, although websites failed to provide comprehensive information about general anesthesia, the information that was included was highly accurate. No website presented any inaccurate information related to items in the coding instrument. However, over half of the websites failed to provide supporting evidence with citations and the credentials of their authors or medical reviewers.

All websites had readability levels above the AMA recommended level of sixth grade at which 99% of US adults can read [16,17]. Websites had a median F-K grade level of 11.3 (IQR 9.5-12.8) and a median SMOG score of 13.5 (IQR 12.2-14.4), consistent with the findings in a previous study about anesthesiology-related patient education materials on the internet [11]. About 90% of US adults attained an education level of high school graduate or higher, equivalent to a readability level of 12, while only 62% attained a grade level of 13 or above [17,71]. Therefore, a considerable portion of the US adult population may have difficulty reading many of the websites on general anesthesia.

Patients report that they are interested in visiting websites about anesthesia if recommended by their clinicians [3-5], so a goal of this study was to identify high-quality websites that can be recommended to patients to aid them in making informed decisions about their anesthetic care. Per the coding instrument used in this study, the websites [46-50] that provided the most comprehensive information about general anesthesia each had its information dispersed over many web pages, so that patients would need to click through linked pages extensively in order to gather all the information. All [48-50] except for Wikipedia.org [46,47] had readability levels below the median values of all websites analyzed. Despite being less comprehensive, ClevelandClinic.org [51,52], MedicalNewsToday.com [53], NHS.uk [54,55], NHSinform.scot [56], and OUH.NHS.uk [57] each condensed its information into ≤3 web pages, allowing for an easier browsing experience. In addition, their readability levels were comparable to, if not lower than, the most comprehensive websites. Overall, those websites that were less comprehensive but easier to browse through, especially the ones [53,54,56,57] that were easier to read, might be more suitable as supplemental patient education recourses that can be recommended to surgical patients.

In terms of websites that provided specific information comparing INVA and TIVA, Wikipedia.org [46,47] and NYSORA.com [63] were the most comprehensive. The former [46,47] had the problems aforementioned, while the latter [63] was highly technical and difficult to read. None of the websites provided comprehensive information with good readability and in a format that made it easy to directly compare INVA and TIVA, highlighting the need to develop patient education materials that address these deficits while summarizing the best evidence currently available. Moreover, consistent with past work [2], the limited availability of information on the internet about INVA and TIVA may reflect a pervasive view or culture that it is not necessary or important to involve patients in decisions surrounding anesthesia care, including the choice of INVA versus TIVA. It may also reflect the fact that there is limited evidence available regarding the comparison between INVA and TIVA, particularly in the setting of noncardiac surgeries, due to the lack of robust comparative effectiveness trials [23,24,72-83]. Future studies are needed to compare patient recovery experiences and outcomes after using INVA versus TIVA for general anesthesia. If a benefit were found to be

XSL•FO
RenderX

associated with 1 method over the other, it would be relevant to patients, and should influence their desire to choose between the 2 general anesthesia administration methods.

Limitations

This study had some limitations. First, although the search terms were designed to reflect both technical terminology and laypersons' language, patients might search with keywords different than the ones used in this study, and obtain different search results. Second, each included web page linked within 2 clicks was considered as a single website for analysis in order to imitate the public's general approach to web browsing. However, in some cases, as many as 20 web pages were included within a single website, whereas it is unlikely for patients to read as extensively to encounter all the information presented. Third, the website selection criteria were designed to capture all websites that patients are likely to encounter when searching for information on general anesthesia; it is possible that patients are aware of the potential for misinformation on the internet and focus instead on a selective subset of websites that they trust. Thus the percentage of higher-quality, more comprehensive websites may be underestimated compared to what patients actually read when they select a more restricted subset. However, the less stringent website selection criteria used in this study conferred a greater ability to identify areas in which information was lacking. Fourth, the keyword searches and the extraction of web page contents were conducted once for each keyword or web page at a single point in time. Website contents could be updated over time, which would not be captured by the cross-sectional approach used in this study. Fifth, all website coding was performed by a single author, so the interrater reliability of the coding instrument cannot be assessed. However, most items in the coding instrument were objective, and any ambiguity encountered during the coding process was discussed among the authors until a consensus was reached. Sixth, videos were excluded from the analysis to ensure the comparability among included websites, but they can be an important source of information requiring future studies to evaluate. Seventh, non-English websites were excluded from the analysis. Evaluating and improving websites and patient education materials written in other languages would be valuable for the large population of non–English-speaking patients. Finally, readability formulas have their limitations. Both formulas used in this study involve assessing the number of syllables [44,45]. The topic word "anesthesia" has 4 syllables and is considered "polysyllabic" per the definition of the SMOG formula [45]. The unavoidable use of such topic words in anesthesiology-related materials may bias the readability scores toward higher values without necessarily creating difficulty for patients to understand. Readability is not a perfect surrogate for comprehensibility, and future studies are needed to assess how well commonly used readability scores correlate with patient comprehension.

Conclusions

Websites about general anesthesia can benefit from additional, more comprehensive information and text readability. While some websites on general anesthesia provided more comprehensive information compared to others, no website on the specific comparison between INVA and TIVA can aid patients in deciding with their clinicians about these anesthetic options. There is a need for high-quality patient education materials about general anesthesia, particularly on INVA versus TIVA, to provide comprehensive, accurate information in a format conducive to patient understanding.

Data Availability

The data sets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

XH, BRTP, NGdB, and MCP conceptualized this study and developed the study methodology. XH conducted investigation, software coding, and formal analysis. BRTP and MCP supervised this study. XH, BRTP, and MCP wrote the original draft of the paper and substantially edited it. MSA and SK reviewed and edited the paper.

Conflicts of Interest

MCP was a consultant for UCB Biopharma in 2022 on a topic unrelated to the content of this paper.

Multimedia Appendix 1
Two-part coding instrument adapted from the International Patient Decision Aids Standards (IPDAS) minimal criteria.
[DOCX File , 24 KB - periop_v6i1e47714_app1.docx]

References

1. Meara JG, Leather AJM, Hagander L, Alkire BC, Alonso N, Ameh EA, et al. Global surgery 2030: evidence and solutions for achieving health, welfare, and economic development. Lancet 2015;386(9993):569-624 [FREE Full text] [doi: 10.1016/S0140-6736(15)60160-X] [Medline: 25924834]
2. Pennington BRT, Politi MC, Abdallah AB, Janda AM, Eshun-Wilsonova I, deBourbon NG, et al. A survey of surgical patients' perspectives and preferences towards general anesthesia techniques and shared-decision making. BMC Anesthesiol 2023;23(1):277 [FREE Full text] [doi: 10.1186/s12871-023-02219-5] [Medline: 37592215]

XSL•FO
RenderX

3. Kurup V, Considine A, Hersey D, Dai F, Senior A, Silverman DG, et al. Role of the internet as an information resource for surgical patients: a survey of 877 patients. Br J Anaesth 2013;110(1):54-58 [FREE Full text] [doi: 10.1093/bja/aes326] [Medline: 22991261]
4. Wieser T, Steurer MP, Steurer M, Dullenkopf A. Factors influencing the level of patients using the internet to gather information before anaesthesia: a single-centre survey of 815 patients in Switzerland: the internet for patient information before anaesthesia. BMC Anesthesiol 2017;17(1):39 [FREE Full text] [doi: 10.1186/s12871-017-0319-1] [Medline: 28270097]
5. Nucci B, Claret PG, Leclerc G, Chaumeron A, Grillo P, Buleon C, et al. Role of the internet as an information resource before anaesthesia consultation: a French prospective multicentre survey. Eur J Anaesthesiol 2017;34(12):831-835 [FREE Full text] [doi: 10.1097/EJA.0000000000000686] [Medline: 28922337]
6. Gustin AN. Shared decision-making. Anesthesiol Clin 2019;37(3):573-580 [FREE Full text] [doi: 10.1016/j.anclin.2019.05.001] [Medline: 31337486]
7. Groves ND, Humphreys HW, Williams AJ, Jones A. Effect of informational internet web pages on patients' decision-making: randomised controlled trial regarding choice of spinal or general anaesthesia for orthopaedic surgery. Anaesthesia 2010;65(3):277-282 [FREE Full text] [doi: 10.1111/j.1365-2044.2009.06211.x] [Medline: 20336817]
8. Tallgren M, Bäcklund M. Patient information about general anaesthesia on the internet. Anaesthesia 2009;64(4):408-415 [FREE Full text] [doi: 10.1111/j.1365-2044.2008.05813.x] [Medline: 19317707]
9. Corcoran TB, Ward M, Jarosz K, Schug SA. The evaluation of anaesthesia-related information on the internet. Anaesth Intensive Care 2009;37(1):79-84 [FREE Full text] [doi: 10.1177/0310057X0903700103] [Medline: 19157351]
10. Marshall R, Pomeroy E, McKendry C, Gilmartin M, McQuail P, Johnson M. Anaesthesia for total hip and knee replacement: a review of patient education materials available online. F1000Res 2019;8:416 [FREE Full text] [doi: 10.12688/f1000research.18675.1] [Medline: 33335711]
11. De Oliveira GS, Jung M, Mccaffery KJ, McCarthy RJ, Wolf MS. Readability evaluation of internet-based patient education materials related to the anesthesiology field. J Clin Anesth 2015;27(5):401-405 [FREE Full text] [doi: 10.1016/j.jclinane.2015.02.005] [Medline: 25912728]
12. Theodosiou CA, Theodosiou LJ. Does the internet provide safe information for pre-anaesthetic patients? Anaesthesia 2003;58(8):805-806 [FREE Full text] [doi: 10.1046/j.1365-2044.2003.03295_4.x] [Medline: 12859480]
13. Caron S, Berton J, Beydon L. Quality of anaesthesia-related information accessed via internet searches. Br J Anaesth 2007;99(2):195-201 [FREE Full text] [doi: 10.1093/bja/aem117] [Medline: 17510047]
14. Harmon D, Duggan M, Flynn N. Anaesthesia on the World Wide Web: is reliable patient information available on the internet? Anaesthesia 2000;55(7):728-729 [FREE Full text] [doi: 10.1046/j.1365-2044.2000.01557-58x./] [Medline: 10919477]
15. Roughead T, Sewell D, Ryerson CJ, Fisher JH, Flexman AM. Internet-based resources frequently provide inaccurate and out-of-date recommendations on preoperative fasting: a systematic review. Anesth Analg 2016;123(6):1463-1468 [FREE Full text] [doi: 10.1213/ANE.0000000000001590] [Medline: 27644057]
16. Weiss BD, Schwartzberg JG. Health Literacy and Patient Safety: Help Patients Understand. Manual for Clinicians. Chicago: American Medical Association Foundation; 2007.
17. Educational attainment in the United States: 2021. United States Census Bureau. 2022. URL: https://www.census.gov/data/tables/2021/demo/educational-attainment/cps-detailed-tables.html [accessed 2022-12-01]
18. Search engine market share worldwide. StatCounter Global Stats. 2022. URL: https://gs.statcounter.com/search-engine-market-share [accessed 2022-03-28]
19. Soltys FC, Spilo K, Politi MC. The content and quality of publicly available information about congenital diaphragmatic hernia: descriptive study. JMIR Pediatr Parent 2021;4(4):e30695 [FREE Full text] [doi: 10.2196/30695] [Medline: 34665147]
20. Madden T, Cortez S, Kuzemchak M, Kaphingst KA, Politi MC. Accuracy of information about the intrauterine device on the internet. Am J Obstet Gynecol 2016;214(4):499.e1-499.e6 [FREE Full text] [doi: 10.1016/j.ajog.2015.10.928] [Medline: 26546848]
21. International Patient Decision Aids Standards (IPDAS) Collaboration. URL: http://ipdas.ohri.ca/using.html [accessed 2021-12-18]
22. Joseph-Williams N, Newcombe R, Politi M, Durand MA, Sivell S, Stacey D, et al. Toward minimum standards for certifying patient decision aids: a modified Delphi consensus process. Med Decis Making 2014;34(6):699-710 [FREE Full text] [doi: 10.1177/0272989X13501721] [Medline: 23963501]
23. Sneyd JR, Carr A, Byrom WD, Bilski AJ. A meta-analysis of nausea and vomiting following maintenance of anaesthesia with propofol or inhalational agents. Eur J Anaesthesiol 1998;15(4):433-445 [FREE Full text] [doi: 10.1046/j.1365-2346.1998.00319.x] [Medline: 9699101]
24. Pandit JJ, Andrade J, Bogod DG, Hitchman JM, Jonker WR, Lucas N, et al. The 5th National Audit Project (NAP5) on accidental awareness during general anaesthesia: summary of main findings and risk factors. Anaesthesia 2014;69(10):1089-1101 [FREE Full text] [doi: 10.1111/anae.12826] [Medline: 25204236]

XSL•FO
RenderX

25. Lauder GR, Thomas M, von Ungern-Sternberg BS, Engelhardt T. Volatiles or TIVA: Which is the standard of care for pediatric airway procedures? A pro-con discussion. Paediatr Anaesth 2020 Mar;30(3):209-220. [doi: 10.1111/pan.13809] [Medline: 31886922]
26. Ellinas H, Albrecht MA. Malignant Hyperthermia Update. Anesthesiol Clin 2020 Mar;38(1):165-181. [doi: 10.1016/j.anclin.2019.10.010] [Medline: 32008650]
27. Hopkins PM, Girard T, Dalay S, Jenkins B, Thacker A, Patteril M, et al. Malignant hyperthermia 2020: Guideline from the Association of Anaesthetists. Anaesthesia 2021 May;76(5):655-664 [FREE Full text] [doi: 10.1111/anae.15317] [Medline: 33399225]
28. Hemphill S, McMenamin L, Bellamy MC, Hopkins PM. Propofol infusion syndrome: a structured literature review and analysis of published case reports. Br J Anaesth 2019 Apr;122(4):448-459 [FREE Full text] [doi: 10.1016/j.bja.2018.12.025] [Medline: 30857601]
29. Gan TJ, Diemunsch P, Habib AS, Kovac A, Kranke P, Meyer TA, et al. Consensus guidelines for the management of postoperative nausea and vomiting. Anesth Analg 2014 Jan;118(1):85-113. [doi: 10.1213/ANE.0000000000000002] [Medline: 24356162]
30. Lewis SR, Pritchard MW, Fawcett LJ, Punjasawadwong Y. Bispectral index for improving intraoperative awareness and early postoperative recovery in adults. Cochrane Database Syst Rev 2019 Sep 26;9(9):CD003843 [FREE Full text] [doi: 10.1002/14651858.CD003843.pub4] [Medline: 31557307]
31. Sebel PS, Bowdle TA, Ghoneim MM, Rampil IJ, Padilla RE, Gan TJ, et al. The incidence of awareness during anesthesia: a multicenter United States study. Anesth Analg 2004 Sep;99(3):833-839. [doi: 10.1213/01.ANE.0000130261.90896.6C] [Medline: 15333419]
32. Practice Guidelines for Preoperative Fasting and the Use of Pharmacologic Agents to Reduce the Risk of Pulmonary Aspiration: Application to Healthy Patients Undergoing Elective Procedures: An Updated Report by the American Society of Anesthesiologists Task Force on Preoperative Fasting and the Use of Pharmacologic Agents to Reduce the Risk of Pulmonary Aspiration. Anesthesiology 2017 Mar;126(3):376-393 [FREE Full text] [doi: 10.1097/ALN.0000000000001452] [Medline: 28045707]
33. Yasny JS, White J. Environmental implications of anesthetic gases. Anesth Prog 2012;59(4):154-158 [FREE Full text] [doi: 10.2344/0003-3006-59.4.154] [Medline: 23241038]
34. Bach A, Motsch J. Infectious risks associated with the use of propofol. Acta Anaesthesiol Scand 1996 Nov;40(10):1189-1196. [doi: 10.1111/j.1399-6576.1996.tb05549.x] [Medline: 8986181]
35. Bennett SN, McNeil MM, Bland LA, Arduino MJ, Villarino ME, Perrotta DM, et al. Postoperative infections traced to contamination of an intravenous anesthetic, propofol. N Engl J Med 1995 Jul 20;333(3):147-154. [doi: 10.1056/NEJM199507203330303] [Medline: 7791816]
36. Edwards A, Thomas R, Williams R, Ellner AL, Brown P, Elwyn G. Presenting risk information to people with diabetes: evaluating effects and preferences for different formats by a web-based randomised controlled trial. Patient Educ Couns 2006 Nov;63(3):336-349. [doi: 10.1016/j.pec.2005.12.016] [Medline: 16860964]
37. Timmermans DRM, Ockhuysen-Vermey CF, Henneman L. Presenting health risk information in different formats: the effect on participants' cognitive and emotional evaluation and decisions. Patient Educ Couns 2008 Dec;73(3):443-447. [doi: 10.1016/j.pec.2008.07.013] [Medline: 18722073]
38. Sheridan SL, Halpern DJ, Viera AJ, Berkman ND, Donahue KE, Crotty K. Interventions for individuals with low health literacy: a systematic review. J Health Commun 2011;16 Suppl 3:30-54. [doi: 10.1080/10810730.2011.604391] [Medline: 21951242]
39. Miller D, Lewis SR, Pritchard MW, Schofield-Robinson OJ, Shelton CL, Alderson P, et al. Intravenous versus inhalational maintenance of anaesthesia for postoperative cognitive outcomes in elderly people undergoing non-cardiac surgery. Cochrane Database Syst Rev 2018 Aug 21;8(8):CD012317 [FREE Full text] [doi: 10.1002/14651858.CD012317.pub2] [Medline: 30129968]
40. Nordström O, Engström AM, Persson S, Sandin R. Incidence of awareness in total i.v. anaesthesia based on propofol, alfentanil and neuromuscular blockade. Acta Anaesthesiol Scand 1997 Sep;41(8):978-984. [doi: 10.1111/j.1399-6576.1997.tb04823.x] [Medline: 9311394]
41. Sandin RH, Enlund G, Samuelsson P, Lennmarken C. Awareness during anaesthesia: a prospective case study. Lancet 2000 Feb 26;355(9205):707-711. [doi: 10.1016/S0140-6736(99)11010-9] [Medline: 10703802]
42. Weiss E, Moore K. An assessment of the quality of information available on the internet about the IUD and the potential impact on contraceptive choices. Contraception 2003;68(5):359-364 [FREE Full text] [doi: 10.1016/j.contraception.2003.07.001] [Medline: 14636940]
43. Readability score. Readable. 2018. URL: https://readable.com/ [accessed 2022-11-07]
44. Kincaid JP, Fishburne RP, Rogers RL, Chissom BS. Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for Navy enlisted Personnel. Defense Technical Information Center. 1975. URL: https://apps.dtic.mil/sti/citations/ADA006655 [accessed 2023-09-28]
45. Mc Laughlin GH. SMOG grading-a new readability formula. J Read 1969:639-646.

XSL•FO
RenderX

46. General anaesthesia. Wikipedia, The Free Encyclopedia. 2022. URL: https://en.wikipedia.org/w/index.php?title=General_anaesthesia&oldid=1128457337 [accessed 2023-01-16]
47. Total intravenous anaesthesia. Wikipedia, The Free Encyclopedia. 2023. URL: https://en.wikipedia.org/w/index.php?title=Total_intravenous_anaesthesia&oldid=1132012476 [accessed 2023-09-28]
48. Whitlock J. How general anesthesia works. verywellhealth. URL: https://www.verywellhealth.com/before-during-and-after-general-anesthesia-4150168 [accessed 2023-01-16]
49. Axtell B. Side effects of general anesthesia: what to expect. Healthline. 2018. URL: https://www.healthline.com/health/side-effects-of-general-anesthesia [accessed 2000-01-17]
50. Propofol generic diprivan: used for anesthesia and sedation. GoodRx. URL: https://www.goodrx.com/propofol [accessed 2023-01-16]
51. Anesthesia. Cleveland Clinic. 2020. URL: https://my.clevelandclinic.org/health/treatments/15286-anesthesia [accessed 2023-01-16]
52. How safe is anesthesia? 5 common concerns. Cleveland Clinic. 2020. URL: https://health.clevelandclinic.org/safe-anesthesia-5-things-know/ [accessed 2023-01-16]
53. Newman T. What to know about general anesthesia. MedicalNewsToday. 2021. URL: https://www.medicalnewstoday.com/articles/265592 [accessed 2023-01-16]
54. General anaesthesia. National Health Service. 2021. URL: https://www.nhs.uk/conditions/general-anaesthesia/ [accessed 2023-01-16]
55. Anaesthesia. National Health Service. 2021. URL: https://www.nhs.uk/conditions/anaesthesia/ [accessed 2023-01-16]
56. General anaesthetic. NHS Inform. 2020. URL: https://www.nhsinform.scot/tests-and-treatments/medicines-and-medical-aids/types-of-medicine/general-anaesthetic [accessed 2023-01-16]
57. Anaesthesia explained: information for patients. Oxford University Hospitals NHS. URL: https://www.ouh.nhs.uk/patient-guide/leaflets/files/13762Panaesthesia.pdf [accessed 2023-01-16]
58. General anesthesia. American Society of Anesthesiologists. URL: https://www.asahq.org/madeforthismoment/anesthesia-101/types-of-anesthesia/general-anesthesia/ [accessed 2023-01-16]
59. General anesthetics. Drugs.com. URL: https://www.drugs.com/drug-class/general-anesthetics.html [accessed 2023-01-16]
60. Durbin K. Propofol. Drugs.com. URL: https://www.drugs.com/propofol.html [accessed 2023-01-16]
61. General anesthesia. Mayo Clinic. 2023. URL: https://www.mayoclinic.org/tests-procedures/anesthesia/about/pac-20384568 [accessed 2023-08-13]
62. McKechnie D. Anaesthesia. Patient. URL: https://patient.info/treatment-medication/anaesthesia [accessed 2023-01-16]
63. Inhaled anesthetics. NYSORA. URL: https://www.nysora.com/anesthesia/inhaled-anesthetics/ [accessed 2023-01-16]
64. Calixte R, Rivera A, Oridota O, Beauchamp W, Camacho-Rivera M. Social and demographic patterns of health-related internet use among adults in the United States: a secondary data analysis of the Health Information National Trends Survey. Int J Environ Res Public Health 2020;17(18):6856 [FREE Full text] [doi: 10.3390/ijerph17186856] [Medline: 32961766]
65. Shinkunas LA, Klipowicz CJ, Carlisle EM. Shared decision making in surgery: a scoping review of patient and surgeon preferences. BMC Med Inform Decis Mak 2020;20(1):190 [FREE Full text] [doi: 10.1186/s12911-020-01211-0] [Medline: 32787950]
66. Ehlers AP, Telem DA. Shared decision-making-it's not for everyone. JAMA Surg 2022;157(5):414 [FREE Full text] [doi: 10.1001/jamasurg.2022.0291] [Medline: 35319750]
67. Baggett ND, Schulz K, Buffington A, Marka N, Hanlon BM, Zimmermann C, et al. Surgeon use of shared decision-making for older adults considering major surgery: a secondary analysis of a randomized clinical trial. JAMA Surg 2022;157(5):406-413 [FREE Full text] [doi: 10.1001/jamasurg.2022.0290] [Medline: 35319737]
68. Légaré F, Witteman HO. Shared decision making: examining key elements and barriers to adoption into routine clinical practice. Health Aff (Millwood) 2013;32(2):276-284. [doi: 10.1377/hlthaff.2012.1078] [Medline: 23381520]
69. Kunneman M, Henselmans I, Gärtner FR, Bomhof-Roordink H, Pieterse AH. Do shared decision-making measures reflect key elements of shared decision making? A content review of coding schemes. Med Decis Making 2019;39(7):886-893 [FREE Full text] [doi: 10.1177/0272989X19874347] [Medline: 31556799]
70. Hoffmann T, Bakhit M, Michaleff Z. Shared decision making and physical therapy: what, when, how, and why? Braz J Phys Ther 2022;26(1):100382 [FREE Full text] [doi: 10.1016/j.bjpt.2021.100382] [Medline: 35063699]
71. Storino A, Castillo-Angeles M, Watkins AA, Vargas C, Mancias JD, Bullock A, et al. Assessing the accuracy and readability of online health information for patients with pancreatic cancer. JAMA Surg 2016;151(9):831-837 [FREE Full text] [doi: 10.1001/jamasurg.2016.0730] [Medline: 27144966]
72. Schraag S, Pradelli L, Alsaleh AJO, Bellone M, Ghetti G, Chung TL, et al. Propofol vs. inhalational agents to maintain general anaesthesia in ambulatory and in-patient surgery: a systematic review and meta-analysis. BMC Anesthesiol 2018;18(1):162 [FREE Full text] [doi: 10.1186/s12871-018-0632-3] [Medline: 30409186]
73. Elbakry AE, Sultan WE, Ibrahim E. A comparison between inhalational (desflurane) and total intravenous anaesthesia (propofol and dexmedetomidine) in improving postoperative recovery for morbidly obese patients undergoing laparoscopic sleeve gastrectomy: a double-blinded randomised controlled trial. J Clin Anesth 2018;45:6-11 [FREE Full text] [doi: 10.1016/j.jclinane.2017.12.001] [Medline: 29223575]

XSL•FO
RenderX

74. Herling SF, Dreijer B, Lam GW, Thomsen T, Møller AM. Total intravenous anaesthesia versus inhalational anaesthesia for adults undergoing transabdominal robotic assisted laparoscopic surgery. Cochrane Database Syst Rev 2017;4(4):CD011387 [FREE Full text] [doi: 10.1002/14651858.CD011387.pub2] [Medline: 28374886]
75. Nimmo AF, Absalom AR, Bagshaw O, Biswas A, Cook TM, Costello A, et al. Guidelines for the safe practice of total intravenous anaesthesia (TIVA): joint guidelines from the Association of Anaesthetists and the Society for Intravenous Anaesthesia. Anaesthesia 2019;74(2):211-224 [FREE Full text] [doi: 10.1111/anae.14428] [Medline: 30378102]
76. Liu T, Gu Y, Chen K, Shen X. Quality of recovery in patients undergoing endoscopic sinus surgery after general anesthesia: total intravenous anesthesia vs desflurane anesthesia. Int Forum Allergy Rhinol 2019;9(3):248-254 [FREE Full text] [doi: 10.1002/alr.22246] [Medline: 30452125]
77. Lee WK, Kim MS, Kang SW, Kim S, Lee JR. Type of anaesthesia and patient quality of recovery: a randomized trial comparing propofol-remifentanil total i.v. anaesthesia with desflurane anaesthesia. Br J Anaesth 2015;114(4):663-668 [FREE Full text] [doi: 10.1093/bja/aeu405] [Medline: 25500679]
78. Zaballos M, Reyes A, Etulain J, Monteserín C, Rodríguez M, Velasco E. Desflurane versus propofol in post-operative quality of recovery of patients undergoing day laparoscopic cholecystectomy. Prospective, comparative, non-inferiority study. Rev Esp Anestesiol Reanim (Engl Ed) 2018;65(2):96-102 [FREE Full text] [doi: 10.1016/j.redar.2017.09.010] [Medline: 29126612]
79. Miller D, Lewis SR, Pritchard MW, Schofield-Robinson OJ, Shelton CL, Alderson P, et al. Intravenous versus inhalational maintenance of anaesthesia for postoperative cognitive outcomes in elderly people undergoing non-cardiac surgery. Cochrane Database Syst Rev 2018;8(8):CD012317 [FREE Full text] [doi: 10.1002/14651858.CD012317.pub2] [Medline: 30129968]
80. Mashour GA, Avidan MS. Intraoperative awareness: controversies and non-controversies. Br J Anaesth 2015;115(Suppl 1):i20-i26 [FREE Full text] [doi: 10.1093/bja/aev034] [Medline: 25735710]
81. Volatile vs total intravenous anaesthesia for major non-cardiac surgery a pragmatic randomised controlled triaL (VITAL). National Institute for Health and Care Research (NIHR). 2021. URL: https://fundingawards.nihr.ac.uk/award/NIHR130573 [accessed 2023-01-20]
82. Volatile anaesthesia and perioperative outcomes related to cancer: the VAPOR-C trial. ClinicalTrials.gov. 2020. URL: https://clinicaltrials.gov/ct2/show/NCT04316013 [accessed 2023-01-20]
83. THRIVE: trajectories of recovery after intravenous propofol vs inhaled VolatilE anesthesia. Patient-Centered Outcomes Research Institute (PCORI). 2021. URL: https://www.pcori.org/research-results/2021/thrive-trajectories-recovery-after-intravenous-propofol-vs-inhaled-volatile-anesthesia [accessed 2023-01-20]

Abbreviations

AMA: American Medical Association
F-K: Flesch-Kincaid
INVA: inhaled volatile anesthesia
IPDAS: International Patient Decision Aids Standards
PONV: postoperative nausea and vomiting
SMOG: Simple Measure of Gobbledygook
TIVA: total intravenous anesthesia

Edited by T de Azevedo Cardoso; submitted 29.03.23; peer-reviewed by M Irwin, YD Cheng; comments to author 09.08.23; revised version received 29.08.23; accepted 19.09.23; published 02.11.23.

Please cite as:
Hu X, Pennington BRT, Avidan MS, Kheterpal S, deBourbon NG, Politi MC
Description of the Content and Quality of Publicly Available Information on the Internet About Inhaled Volatile Anesthesia and Total Intravenous Anesthesia: Descriptive Study
JMIR Perioper Med 2023;6:e47714
URL: https://periop.jmir.org/2023/1/e47714
doi:10.2196/47714
PMID:37917148

©Xinwen Hu, Bethany R Tellor Pennington, Michael S Avidan, Sachin Kheterpal, Nastassjia G deBourbon, Mary C Politi. Originally published in JMIR Perioperative Medicine (http://periop.jmir.org), 02.11.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Perioperative

XSL•FO
RenderX

Medicine, is properly cited. The complete bibliographic information, a link to the original publication on http://periop.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Perioperative Risk Assessment of Patients Using the MyRISK Digital Score Completed Before the Preanesthetic Consultation: Prospective Observational Study

Fabrice Ferré[1*], MD, PhD; Rodolphe Laurent[1*], MD; Philippine Furelau[1], MD; Emmanuel Doumard[2], PhD; Anne Ferrier[1], MD; Laetitia Bosch[1], MD; Cyndie Ba[1], MD; Rémi Menut[1], MD; Matt Kurrek[3], MD; Thomas Geeraerts[1], MD, PhD; Antoine Piau[4], MD, PhD; Vincent Minville[1], MD, PhD

[1]Département d'Anesthésie-Réanimation, Hôpital Pierre-Paul Riquet, Centre Hospitalier Universitaire Purpan, Toulouse, France

[2]Institut de Recherche en Informatique de Toulouse, Université Toulouse III Paul Sabatier, Toulouse, France

[3]Department of Anesthesia, University of Toronto, Toronto, ON, Canada

[4]Département de Gériatrie, Centre Hospitalier Universitaire Rangueil, Toulouse, France

[*]these authors contributed equally

Corresponding Author:
Fabrice Ferré, MD, PhD
Département d'Anesthésie-Réanimation
Hôpital Pierre-Paul Riquet
Centre Hospitalier Universitaire Purpan
Place du Dr Baylac
Toulouse, 31300
France
Phone: 33 0561779988
Fax: 33 0561772170
Email: fabriceferre31@gmail.com

Abstract

Background: The ongoing COVID-19 pandemic has highlighted the potential of digital health solutions to adapt the organization of care in a crisis context.

Objective: Our aim was to describe the relationship between the *MyRISK* score, derived from self-reported data collected by a chatbot before the preanesthetic consultation, and the occurrence of postoperative complications.

Methods: This was a single-center prospective observational study that included 401 patients. The 16 items composing the MyRISK score were selected using the Delphi method. An algorithm was used to stratify patients with low (*green*), intermediate (*orange*), and high (*red*) risk. The primary end point concerned postoperative complications occurring in the first 6 months after surgery (composite criterion), collected by telephone and by consulting the electronic medical database. A logistic regression analysis was carried out to identify the explanatory variables associated with the complications. A machine learning model was trained to predict the MyRISK score using a larger data set of 1823 patients classified as *green* or *red* to reclassify individuals classified as *orange* as either *modified green* or *modified red*. User satisfaction and usability were assessed.

Results: Of the 389 patients analyzed for the primary end point, 16 (4.1%) experienced a postoperative complication. A *red* score was independently associated with postoperative complications (odds ratio 5.9, 95% CI 1.5-22.3; *P*=.009). A *modified red* score was strongly correlated with postoperative complications (odds ratio 21.8, 95% CI 2.8-171.5; *P*=.003) and predicted postoperative complications with high sensitivity (94%) and high negative predictive value (99%) but with low specificity (49%) and very low positive predictive value (7%; area under the receiver operating characteristic curve=0.71). Patient satisfaction numeric rating scale and system usability scale median scores were 8.0 (IQR 7.0-9.0) out of 10 and 90.0 (IQR 82.5-95.0) out of 100, respectively.

Conclusions: The MyRISK digital perioperative risk score established before the preanesthetic consultation was independently associated with the occurrence of postoperative complications. Its negative predictive strength was increased using a machine learning model to reclassify patients identified as being at intermediate risk. This reliable numerical categorization could be used to objectively refer patients with low risk to teleconsultation.

XSL•FO
RenderX

(JMIR Perioper Med 2023;6:e39044) doi:10.2196/39044

KEYWORDS

chatbot; digital health; preanesthetic consultation; perioperative risk; machine learning; mobile phone

Introduction

Background

In France, the process of a patient undergoing elective surgery includes several essential steps such as the surgical consultation, preanesthetic consultation (PAC), and preanesthetic visit [1]. The decree of December 5, 1994, explicitly states that an anesthesiologist should carry out the PAC [2]. This consultation contributes to the preanesthetic evaluation of the patient's health status, justifying the prescription of complementary examinations (eg, laboratory tests) and any specialized consultations that allow a perioperative risk assessment formalized by the American Society of Anesthesiologists (ASA) score [3]. This perioperative risk evaluation is used, for example, to determine a patient's eligibility for an ambulatory care pathway, an enhanced recovery after surgery program or, conversely, a postoperative stay in the intensive care unit [4,5]. As it stands, this state-of-the-art evaluation requires medical expertise.

Beyond the low reproducibility of the ASA score [6,7], the perioperative risk global assessment during the PAC is not well standardized and may be incomplete, especially when consultation time is limited. This is why in some anesthesia teams, patients are asked to complete a questionnaire in paper form before their PAC, allowing them to specify, for example, their past medical and surgical history or their usual treatments. The patient is then asked to hand the completed questionnaire to the anesthetist during the PAC [8]. However, electronic patient-reported outcome measures offer many advantages over paper-based collection [9-13]: preferred modality; (directly) visualized results; higher data quality and response rate; decreased completion time; facilitates patient-clinician communication, improving the decision-making process; and so on. With regard to the preanesthetic questionnaire, the digital version is considered more efficient than the paper form [14]. Moreover, it has been shown that the quality of perioperative care can be improved by a digitalized preoperative information and assessment program [14-16], particularly through automatic reminders or clinical decision support. Health digital tools are therefore definitely of major interest in the perioperative setting.

Objectives

In 2020, the COVID-19–related restrictions accelerated the implementation of organizational digital health innovations. In the context of the COVID-19 crisis, the anesthesia department at the Toulouse University Hospital in Purpan, Toulouse, France, decided to digitalize the PAC by implementing teleconsultations (*as much as possible* to reduce interpersonal contact) and a digital conversational agent (aka chatbot) that allowed collection of medical data before the PAC. An approach assessing the relevance of the data collected as well as user satisfaction seemed essential to validate the sustainable use of this digital tool. The main objective of this study was to demonstrate that our chatbot was able to stratify patients according to their perioperative risk level. We hypothesized that the *MyRISK* perioperative risk score, established before the PAC according to a predefined algorithm based on data collected digitally, was correlated with the occurrence of postoperative complications at 6 months. Our secondary objectives were to improve the prognostic predictive value of this score using a machine learning model to reclassify patients classified as intermediate risk and assess patients' and physicians' satisfaction when using this digital health tool.

Methods

Experimental Design

This single-center prospective observational study was conducted in the anesthesia department of the Toulouse University Hospital. To our knowledge, the correlation with postoperative complications of a digital score based on self-reported medical data has never been described. Thus, no assumptions could be made regarding the relative risk and the positive and negative predictive values of being classified as high perioperative risk by the MyRISK score. Given the relatively low postoperative complication rate after scheduled orthopedic surgery (almost 5% [17]), we estimated that approximately 500 patients should be included to meet our objectives (based on expert opinion).

Population

All patients aged >18 years who were scheduled for orthopedic surgery at the Toulouse University Hospital between June 1, 2020, and October 31, 2020, were eligible. The exclusion criteria were the presence of a protection regime for adults (guardianship, curatorship, or safeguard of justice), patients who did not speak French, the presence of a major sensory handicap (blindness or deafness) compromising the comprehension of the information, patients who did not complete the digital questionnaire through the chatbot (this criterion was considered a refusal to participate), and patients who expressed their opposition to participating in this study.

MyRISK Score

The preanesthetic digital conversational agent, *Medical Assistant Experience* (*MAX*), was developed by 2 anesthetists of the Toulouse University Hospital (FF and VM) in collaboration with a company that creates secure health companions (BOTdesign, Toulouse, France; Figure 1; Multimedia Appendices 1 and 2). Its content was based on the preexisting paper form questionnaire with the addition of anesthetic items considered relevant, such as those allowing the calculation of perioperative scores published in the literature. As an example, we can cite the calculation of the Amsterdam Preoperative Anxiety and Information Scale score [18]; the Lee cardiovascular complication risk score [19]; or the snoring, tiredness, observed apnea, blood pressure, BMI, age, neck

XSL•FO
RenderX

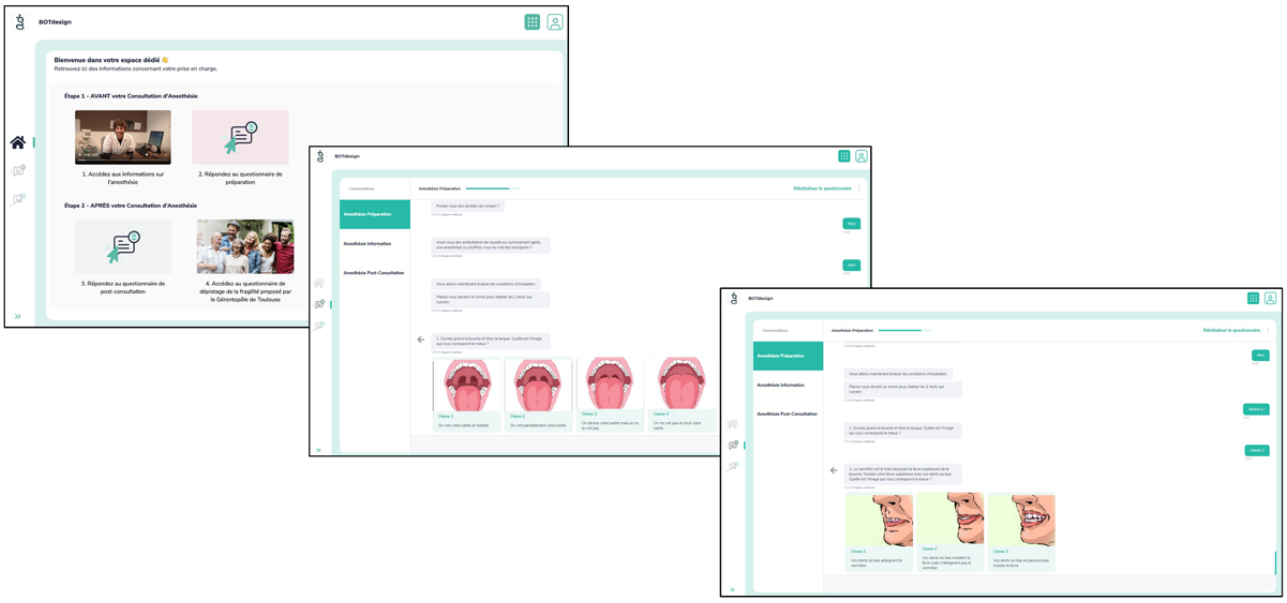
circumference, and gender (STOP-BANG) obstructive sleep apnea screening score [20] in which 7 of the 8 items are collected from the patient's responses to the conversational agent.

Access to the chatbot was made possible once the surgical decision was made, after which the patient received an email inviting them to create their personal account using a smartphone, tablet device, or computer. The data collected were editable at any time by the patient.

The MyRISK score was developed using the Delphi method [21]. The first step was to identify among all the items of data collected by the chatbot those that were relevant, that is, considered to have weight in the development of a predictive risk score. After 2 rounds of discussion, a panel of 6 experts reached a consensus on 16 items, which were then retained (Table 1). The second step involved defining the independent risk level (1, 2, or 3) of each of the 16 items (Table 1). The third step concerned developing an algorithm based on these 16 items to stratify the global perioperative risk level into 3 categories corresponding to a presumed low, intermediate, or high global perioperative risk. Briefly, patients were classified as low risk when all 16 criteria were level 1, as intermediate risk when ≥1 of the 16 criteria were level 2, and as high risk when ≥3 level 2 criteria or ≥1 level 3 criterion were present.

Finally, to make the MyRISK score a visual tool, a *green*, *orange*, or *red* dot was assigned to the low, intermediate, and high global perioperative risk levels, respectively. This color coding was easily accessible and visible on the digital dashboard of patients enrolled in the MAX program. Patients were then considered to have a *green*, *orange*, or *red* MyRISK score.

Figure 1. Screenshots of the digital conversational agent Medical Assistant Experience (BOTdesign, Toulouse, France). The patient is asked to complete a self-assessment of its predictive criteria for difficult intubation (Mallampati and upper lip bite tests).



XSL•FO
RenderX

Table 1. MyRISK score criteria.

Criteria	Level 1	Level 2	Level 3
Age (years) [22]	<65	65 to 80	>80
BMI (kg/m^2) [23]	<30	30 to 40	>40
Drug allergies	No	Yes	N/A[a]
Hemostasis disorders	No	Yes	N/A
Number of medications	0	1 to 5	>5
Active smoking [24]	No	Yes	N/A
Asthma	No	Yes	N/A
Sleep apnea syndrome [25]	No	Yes	N/A
Maximum level of activity (MET[b]) [26]	Walking up 2 flights of stairs without stopping; walking in the street (5-7 km/h); important domestic activities (washing the floor); and sports activities	N/A	Activities of daily living (meals and toileting); walking in the house; and walking in the street (3-5 km/h)
Cardiac symptoms during exercise	No	N/A	Yes
Hypertension	No	Yes	N/A
Cardiac disease [27]	No	Yes	N/A
Respiratory disease [24]	No	Yes	N/A
Renal disease [28]	No	Yes	N/A
Neurological disorders [29]	No	Yes	N/A
Diabetes	No	Yes	N/A

[a]N/A: not applicable.

[b]MET: metabolic equivalent.

Postoperative Complications

Each patient was interviewed by telephone 6 months after surgery by one of the physicians involved in the study. After information was provided on study objectives and oral consent, the telephone survey was used to ask patients about the potential occurrence of postoperative complications. The survey was guided by a computerized structured questionnaire, allowing the secure collection of pseudonymized data. After oral consent, the computerized postoperative patient record form was consulted to ensure that there were no missing data concerning the occurrence of postoperative complications. The average duration of the interview was 9.8 (SD 9) minutes.

We considered that patients had a postoperative complication if they had experienced at least one adverse event among those listed in Textbox 1.

XSL•FO
RenderX

Textbox 1. List of adverse events considered.

Potential postoperative complications

- Acute renal failure (creatinine increase ≥0.3 mg/dL [≥26.5 µmol/L] in 48 hours or creatinine increase ≥1.5×baseline creatinine in <7 days or diuresis <0.5 mL/kg/h for 6 hours or hospitalization for acute kidney injury) [30]
- Myocardial infarction (such as an increase in troponin associated with at least one of the following: signs of ischemia, ST-segment change, development of left branch block on electrocardiogram, or hospitalization for angina or myocardial infarction) [31-33]
- Acute heart failure (such as the presence of clinical, radiological, or echocardiographic signs; N-terminal pro brain natriuretic peptide elevation ≥900 pg/mL; or hospitalization for cardiac heart failure or cardiogenic pulmonary edema) [31,32]
- De novo atrial fibrillation (confirmed on electrocardiogram) [31,34]
- Transient ischemic attack or stroke (confirmed by computed tomography or magnetic resonance imaging or any hospitalization for stroke or transient ischemic attack) [35,36]
- Infection (such as fever requiring antibiotic therapy, hospitalization for fever, or suspected or documented infection)
- Respiratory complication (such as lung disease or respiratory compromise requiring oxygen or noninvasive or invasive ventilation or any hospitalization for respiratory compromise or lung disease) [37]
- Thromboembolic event (confirmed on Doppler ultrasound or computed tomography) [38]
- Hemorrhage (requiring transfusion)
- Rehospitalization
- Death

User Satisfaction

Patient satisfaction with the quality and usability of the chatbot was assessed by the system usability scale (SUS). The SUS is a validated standardized tool for collecting users' opinions on the perceived ease of use of a digitalized system [39,40].

Briefly, the SUS assesses user experience and acceptability (Table 2). Ten statements (5 positive and 5 negative) are listed. Users respond to each statement using a Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). The overall score is calculated to consider items with reversed valences. The final score is between 0 and 100; a score between 50 and 75 is considered fair, 75 to 85 is considered good, and >85 is considered excellent.

Patients' overall satisfaction with the use of the digital conversational agent and with the course of the PAC (ie, face-to-face consultation or teleconsultation) was collected via a simple numeric rating scale (NRS) ranging from 0 (extremely dissatisfied) to 10 (extremely satisfied).

The anesthesiologists using the digital platform during this period were asked to assess their level of satisfaction by using the same evaluation scales (NRS and SUS).

Table 2. System Usability Scale (standard English version).

Statements	Scoring[a]				
	1	2	3	4	5
I think that I would like to use this system					
I found the system unnecessarily complex					
I thought the system was easy to use					
I think that I would need the support of a technical person to be able to use this system					
I found the various functions in the system were well integrated					
I thought there was too much inconsistency in this system					
I would imagine that most people would learn to use this system very quickly					
I found the system very cumbersome to use					
I felt very confident using the system					
I needed to learn a lot of things before I could get going with this system					

[a]Instruction on using the System Usability Scale: Please, circle the appropriate score for each statement from 1 (strongly disagree) to 5 (strongly agree).

Machine Learning Model

Our goal was to train a machine learning model to predict the MyRISK score of patients having either a *green* or *red* score. This trained model was then asked to predict the MyRISK score of individuals classified as *orange* and reclassify them as either *modified green* or *modified red*.

XSL•FO
RenderX

Data Preprocessing

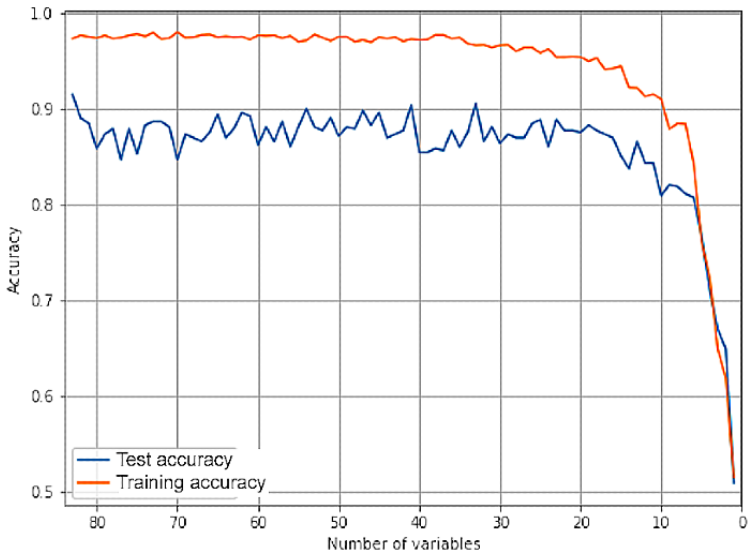
The data set used to train the model was extracted from a larger database of patients. We filtered out duplicates of individuals and features with >70% missing values. We also transformed nominal features following the *one-hot encoding* method, creating a new binary feature for each unique value. We finally filtered out individuals classified as *orange* for the prediction task. The final processed data set was composed of 1823 individuals classified as *green* or *red* for 83 features.

Feature Selection

To filter out redundant features, the recursive feature elimination (RFE) method was used [41]. Briefly, the following steps were applied: data were split into training and test sets; the model was trained on the training set, and its performance was evaluated on the test set; each feature contribution was evaluated, and the least contributing feature was identified (local Shapley additive explanations method [42]) and removed before going back to the first step.

The RFE algorithm returns a list of model performances in the training and test sets for each feature. The evolution of the training and test accuracies of the model through the RFE are presented in Figure 2. In total, 25 features were finally selected in the data set for the final model training.

Figure 2. Evolution of the training and test accuracies of the model through the recursive feature elimination method. The test performance of the model starts to worsen significantly when <25 features are considered. The 25 most contributing features were finally selected for the final model training.



Model Training and Performance

The extreme gradient boosting classifying model was used to train the model on the 25 selected features [43]. Using a hyperparameter grid search, the training and test confusion matrices were obtained with good accuracy scores (97.5% and 96%, respectively).

All processing stages were performed using open-source software machine learning libraries sklearn 1.0.1 and extreme gradient boosting 1.5 (for the model) in Python programming language (version 3.9.7; Python Software Foundation).

Statistical Analysis

Qualitative data were expressed as numbers (%). Quantitative data were expressed as median (IQR) or mean (SD) as appropriate. Categorical variables were compared using the Fisher exact test or the chi-square test. Quantitative variables were compared using the nonparametric Mann-Whitney *U* test. Multivariate analysis (stepwise logistic regression analysis) was performed to identify explanatory variables for the occurrence of postoperative complications at 6 months. The analysis was performed on an intention-to-treat basis. Statistical analysis was performed using MedCalc (version 12.6.1; MedCalc Software Ltd). $P<.05$ was considered statistically significant.

Ethics Approval

This research is considered an experiment in educational sciences aiming to evaluate the participative and pedagogical quality of a new digital tool implemented in current practice. Hence, this research was deemed to fall outside the Jardé law, meaning that no formal ethics approval was required for this study.

Information about the participants' health conditions was collected by the chatbot after they had created their account, but the company BOTdesign had access neither to the patients' identity nor to their IP address. This strategy of data protection was decided in agreement with the eHealth committee of the University Hospital Center of Toulouse. The connection to the digital questionnaire was secure (following the *General Data Protection Regulation* guidelines). An email invitation to log in was sent to the patients after the appointment with the surgeon. Patients chose their own secret password to create their MAX account. Each patient was given oral and written information about this research before enrollment, ensuring that they did not have any objection to participating. This study did not present any risk for the participants and did not modify the usual care pathway or the time required for patient management.

XSL•FO
RenderX

Results

Population

Of the 1000 eligible patients who were scheduled for orthopedic surgery at the Toulouse University Hospital between June 1, 2020, and October 31, 2020, a total of 434 (43.4%) patients logged in to the chatbot. Of these 434 patients, 401 (92.4%) agreed to participate in this study. The characteristics of the studied population are presented in Table 3.

Table 3. Demographic characteristics of the patients.

Characteristics	Values
Age (years), median (IQR)	39 (27-54)
Sex (n=401), n (%)	
Male	241 (60.1)
Female	160 (39.9)
BMI (kg/m^2), mean (SD)	25 (4)
Education (n=388), n (%)	
Undergraduate	83 (21.4)
Graduate	135 (34.8)
Postgraduate	170 (43.8)
Ambulatory care pathway (n=388), n (%)	297 (76.5)
Surgical risk (n=389), n (%)	
Minor	284 (73)
Intermediate	89 (22.9)
Major	16 (4.1)
ASA[a] score (n=389), n (%)	
1	282 (72.5)
2	89 (22.9)
3	18 (4.6)
4	0 (0)
MyRISK score (n=389), n (%)	
Green (low risk)	100 (25.7)
Orange (intermediate risk)	150 (38.6)
Red (high risk)	139 (35.7)
STOP-BANG[b] score modified (out of 7), median (IQR)	1 (1-2)
Lee score modified (out of 4), median (IQR)	0 (0-1)
APAIS[c] anesthesia score (out of 15), median (IQR)	5 (3-7)

[a]ASA: American Society of Anesthesiologists.

[b]STOP-BANG: snoring, tiredness, observed apnea, blood pressure, BMI, age, neck circumference, and gender.

[c]APAIS: Amsterdam Preoperative Anxiety and Information Scale.

User Satisfaction

The median satisfaction score of patients regarding the use of the chatbot as assessed by the NRS was 8.0 (IQR 7.0-9.0) out of 10. The median satisfaction score regarding the use of the digital questionnaire as assessed by the SUS was 90.0 (IQR 82.5-95.0) out of 100.

The median SUS score was higher for users who chose a tablet device or smartphone (97/391, 24.8%) than for those who chose a computer (294/391, 75.2%): 92.5 (IQR 85.0-97.5) versus 90.0 (IQR 82.0-95.0), respectively (*P*=.01).

A large majority of the PACs were teleconsultations (331/401, 82.5%). Regarding the patients' wishes for a future PAC if indicated, 54.7% (181/331) of the patients who received a teleconsultation wished to keep this mode of PAC in the future, whereas 19% (13/70) of the patients who received a face-to-face consultation wished to keep the same mode of PAC (*P*=.08). The mean patient satisfaction score (NRS) regarding the teleconsultation was 8.4 (SD 1.59) out of 10.

XSL•FO
RenderX

The satisfaction score of the anesthesiologists (n=18) regarding the use of the digital platform was collected. Their median satisfaction score was 7.0 (IQR 6.0-8.0) out of 10, and their median SUS usability score was 72.5 (IQR 63.1-88.1) out of 100.

Postoperative Complications

Of the 389 patients analyzed, 16 (4.1%) had a postoperative complication at 6 months. No deaths were reported. The results of the univariate analysis are presented in Table 4.

A dependency relationship between the ASA and MyRISK scores was found (Table 5).

Compared with ASA score=1, an ASA score of ≥3 was independently associated with the occurrence of postoperative complications at 6 months (odds ratio [OR] 5.8, 95% CI 1.7-20.2; *P*=.006). In comparison with a *green* score, a *red* score was independently associated with the occurrence of postoperative complications at 6 months (OR 5.9, 95% CI 1.5-22.3; *P*=.009). Age and surgical risk included in the analysis were not identified as independent variables of the occurrence of postoperative complications. The area under the receiver operating characteristic curve (AUC) of the selected model was 0.78 (95% CI 0.73-0.82).

Finally, a *red* score predicted postoperative complications with 75% sensitivity, 98% negative predictive value, 66% specificity, and 9% positive predictive value (AUC=0.70).

XSL•FO
RenderX

Table 4. Comparison between patients with postoperative complications and those without postoperative complications (univariate analysis; N=389).

	No postoperative complications (n=373)	Postoperative complications (n=16)	*P* value
Age (years), median (IQR)	39.0 (27.0-53.0)	56.5 (44.0-68.0)	.007
Ambulatory care pathway, n (%)	292 (78.3)	5 (31.2)	<.001
Surgical risk, n (%)			.06
Minor	276 (74)	8 (50)	
Intermediate	83 (22.2)	6 (37.5)	
Major	14 (3.8)	2 (12.5)	
ASA[a] score, n (%)			<.001
1	277 (74.3)	5 (31.2)	
2	83 (22.2)	6 (37.5)	
3	13 (3.5)	5 (31.2)	
4	0 (0)	0 (0)	
MyRISK score, n (%)			.002
Green (low risk)	100 (26.8)	0 (0)	
Orange (intermediate risk)	146 (39.1)	4 (25)	
Red (high risk)	127 (34.1)	12 (75)	
Number of medications, median (IQR)	0.0 (0.0-2.0)	0.5 (0.0-1.0)	.65
Active smoking, n (%)	172 (46)	8 (50)	.80
Asthma, n (%)	43 (11.5)	2 (12.5)	.99
Sleep apnea syndrome, n (%)	14 (3.7)	2 (12.5)	.14
Cardiovascular disease, n (%)	24 (6.4)	3 (18.7)	.09
Renal disease, n (%)	7 (1.8)	0 (0)	.99
Neurological disease, n (%)	26 (6.9)	2 (12.5)	.33
Digestive disease, n (%)	55 (14.7)	3 (18.7)	.72
Diabetes, n (%)	19 (5)	1 (6.2)	.58
APAIS[b] anesthesia score (out of 15), median (IQR)	5.0 (3.0-7.0)	6.0 (4.0-7.5)	.61
Lee score modified (out of 4), median (IQR)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	.52
Apfel score modified (out of 3), median (IQR)	1.0 (0.0-1.0)	1.0 (0.0-2.0)	.48
STOP-BANG[c] score modified (out of 7), median (IQR)	1.0 (1.0-2.0)	2.0 (1.0-2.5)	.46

[a]ASA: American Society of Anesthesiologists.

[b]APAIS: Amsterdam Preoperative Anxiety and Information Scale.

[c]STOP-BANG: snoring, tiredness, observed apnea, blood pressure, BMI, age, neck circumference, and gender.

Table 5. Correlation between American Society of Anesthesiologists (ASA) and MyRISK scores (N=400).

MyRISK score	ASA score, n (%)			*P* value
	1	2	3	
Green (low risk)	98 (24.4)	7 (1.7)	0 (0)	<.001
Orange (intermediate risk)	124 (30.9)	28 (7)	1 (0.2)	<.001
Red (high risk)	66 (16.5)	58 (14.5)	18 (4.5)	<.001

XSL•FO
RenderX

Recalculation of the Predictive Value of the MyRISK Score Using a Machine Learning Model

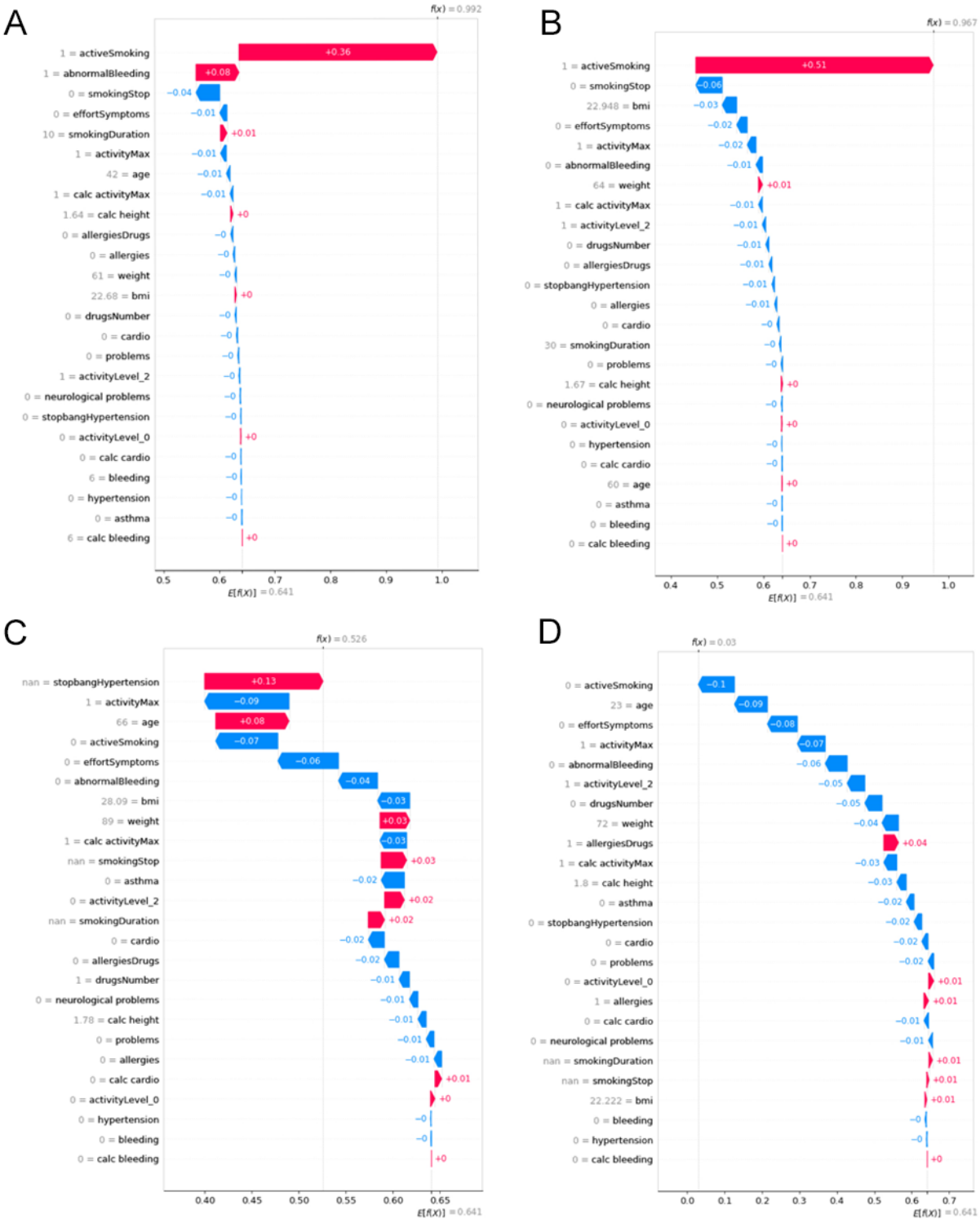
Reclassification of Patients With an Orange MyRISK Score

Among the 389 patients analyzed for the primary end point, 150 (38.6%) were initially classified as *orange*. Of these 150 patients, 4 (2.7%) experienced postoperative complications. Of the 146 patients classified as *orange* with no postoperative complications, 65 (44.5%) were reclassified as *modified red* and 81 (55.5%) as *modified green* using the trained model. Similarly, of the 4 patients with postoperative complications, 3 (75%) were finally reclassified as *modified red* and 1 (25%) as *modified green*.

Concerning these 4 patients, the contribution of each feature was computed with the local Shapley additive explanations method (refer to the Methods section for details). The results are presented in Figure 3.

XSL•FO
RenderX

Figure 3. Computation of the contribution of each of the 25 features (local Shapley additive explanations method) of the 4 patients with an orange MyRISK score who experienced postoperative complications. The first 3 patients (represented by A, B, and C) were reclassified as modified red by the machine learning model; (D) represents the patient who was finally predicted modified green.



Predictive Value of the Modified MyRISK Score

Once the 4 patients classified as *orange* were reclassified, a *modified red* MyRISK score was identified as strongly associated with postoperative complications at 6 months (OR 21.8, 95% CI 2.8-171.5; *P*=.003). An ASA score of ≥3 was also associated with postoperative complications (OR 4.7, 95% CI 1.4-16; *P*=.01).

Finally, a *modified red* score predicted postoperative complications with high sensitivity (94%) and high negative predictive value (99%) but with low specificity (49%) and very low positive predictive value (7%; AUC=0.71).

XSL•FO
RenderX

Discussion

Principal Findings

Through this study, we were able to validate the prognostic predictive value of a perioperative risk score established from data collected by a digital conversational agent implemented to assist the anesthetist during the PAC. In this context, the MyRISK score was correlated with the incidence of complications occurring in the first 6 months postoperatively. The strength of its predictive value was increased using a machine learning model to reclassify patients classified as intermediate risk. The use of an objective method allowing perioperative risk stratification according to a color code (ie, visual tool) and available before the PAC could be relevant for physicians. Finally, the use of this innovative digital tool seems to fully satisfy users.

We were able to identify a dependency between the level of perioperative risk at the end of the medical clinical evaluation (ie, ASA score) and the one calculated digitally before the PAC (ie, MyRISK score). This correlation had already been found by Zuidema et al [44] in a 2011 study using a 22-item numerical questionnaire administered to 14,349 patients, the authors highlighted that a computerized risk assessment could perform well and correlate with the clinical assessment of the ASA score (AUC=0.953), while limiting the interindividual variability of a clinician-assessed ASA score. In this context, it is interesting to note that the primary factor in numerical misclassification of the ASA score was an incomplete or incorrect patient response. More recently, Enneking et al [45] presented a 5-criteria composite preoperative risk score (*patient-centered anesthesia triage system score*) that correlated well with the ASA score (AUC=0.75, 95% CI 0.69-0.83), highlighting its usefulness for patient triage. Since 2017, the authors have been using this score in clinical practice to propose the systematic performance of a teleconsultation for patients classified as no risk.

By analyzing complications occurring in the first 6 postoperative months, we were able to validate the independent prognostic predictive value of the MyRISK score on the occurrence of serious postoperative adverse events. To our knowledge, this approach of validating a numerical risk score on objective criteria (ie, postoperative complications) has never been described in the literature. Thus, by reliably classifying patients according to their level of perioperative risk, the use of the (modified) MyRISK score could allow, before the PAC, triaging of patients by proposing the most appropriate modality of consultation (eg, teleconsultation for patients classified as [modified] green and face-to-face consultation for patients classified as [modified] red). Secondary benefits linked to this triage modality are expected: patients' experience could be improved by reducing waiting time and optimizing consultation time. In addition, the face-to-face consultation could be dedicated to the management of patients with the most complex conditions who require, for example, specialized examinations or consultations. Moreover, as the role of anesthesiologists in the postoperative management of patients is growing, the implementation of a postanesthetic consultation could be a future trend, particularly for patients classified as high risk. In this setting, digital tools help to keep patients and caregivers connected for better follow-up, allowing an early detection or even prevention of postoperative complications.

The current health context has temporarily established the need to reduce travel throughout the country to limit interpersonal contact. Thus, the French National Authority for Health has recommended the use of telemedicine to enable remote management of patients [46]. Finally, the COVID-19 pandemic has highlighted the potential of digital health solutions in facilitating the adaptation of the organization of care in a crisis situation [47]. In this context, the development of the digital solution *MAX* helped us to organize the resumption of surgical activity when restrictions were lifted. A relevant distribution of patients between teleconsultation and face-to-face consultation using the (modified) MyRISK score could allow the indefinite continuance of teleconsultation after the pandemic. However, only 25.7% (100/389) of the patients were classified as low perioperative risk (ie, *green*). Questions still remain as to the proper organizational management of patients classified as intermediate risk (ie, *orange*). It is worth noting that the percentage of patients classified as low perioperative risk (theoretically eligible for teleconsultation) increases to 47% (183/389) when using the modified classification.

In our study, patient satisfaction with the use of the digital questionnaire as well as the PAC process was excellent. The developed digital conversational agent seems to be an adequate platform for the collection of patients' medical information, as shown by the excellent usability score obtained. Moreover, the usability seems to be better when completing the digital questionnaire on a smartphone or tablet device. Our results highlight the enthusiasm of patients for using a digital health platform. These results are in agreement with those described by VanDenKerkhof et al [14], where patient comfort was increased by >70% by computerizing the preanesthetic questionnaire.

Although we did not strictly evaluate the reliability of the data collected by the digital conversational agent, Osman et al [48] demonstrated a response reliability of >90% when a computerized preanesthetic questionnaire was used. Thus, there is consensus in the literature now of the reliability of digital collection of information [49].

Several factors may have favored patient acceptance of this new digital solution. The young age of the patients enrolled (median age 39.0, IQR 27.0-54.0 years) and their level of education (305/389, 78.4%, had graduate or postgraduate degrees) probably explain the very high levels of satisfaction and usability obtained. These results are in agreement with those obtained by Kruse et al [50]. Indeed, the age, level of education, and computer skills of the patients were the 3 main barriers to telehealth adoption identified by the authors [50]. The acceptability and satisfaction of the patients obtained during the use of this digital support encourages our department to develop telemedicine solutions. This enthusiasm is reinforced by the good satisfaction ratings provided by the members of the medical team during the use of this platform, which is probably linked to the automatic integration of the data collected by the digital conversational agent into the computerized PAC file.

XSL•FO
RenderX

However, a qualitative analysis of the main difficulties encountered by the physicians highlighted the absolute necessity of good interoperability among the various software systems.

In our study, the rate of postoperative complications observed at 6 months was 4.1% (16/389). No deaths were recorded. These results are in accordance with those already published in the literature. Indeed, in 2013, Chikuda et al [51] evaluated postoperative morbidity and mortality rates in scheduled orthopedic surgery among >100,000 patients [51]. In this context, the morbidity and mortality rates were 4.2% and 0.11%, respectively.

Our study includes several limitations. First, the results obtained in preoperative scheduled surgeries cannot be extrapolated to the context of urgent surgeries where the use of a digital conversational agent to assist the PAC seems difficult to achieve. Second, 82.5% (331/401) of the PACs analyzed in the study were teleconsultations. The period when the patients were included corresponded to the end of the first lockdown, which explains this high rate of teleconsultations that is not very representative of the subsequent evolution of the practices of our unit (approximately 40% currently). Third, the potential benefits of allocating patients to teleconsultation or face-to-face consultation according to the MyRISK score deserve to be studied in more detail; for example, analysis of patient-perceived quality-of-care indicators (eg, patient-reported outcome measures and patient-reported experience measures) [52] could demonstrate that the experience of care perceived by patients classified as *green* receiving teleconsultation is optimal, notably by avoiding unnecessary travel that disrupts their personal and professional schedules. Fourth, collection of the overall consumption of care by consulting the national health data system could have increased the prognostic predictive value of the MyRISK score by a more global and exhaustive analysis of the postoperative evolution of patients. This type of medicoeconomic approach should be favored in the future. Fifth, the experimental design of our study did not allow us to evaluate the reasons for the nonconnection to MAX by a significant number of patients. Thus, it seems likely that the satisfaction and usability scores were overestimated because they were collected only from the user population of patients. Analysis of the overall data is fundamental to understanding the potential explanatory factors of the digital divide in this context. Fifth, further studies are needed to extend the validation of the MyRISK score to other surgical populations and thus generalize our results. Finally, there is a need for studies examining the impact that the use of these tools has on clinical decision-making and patient outcomes.

Conclusions

To conclude, we were able to demonstrate the prognostic value of a perioperative risk score established from data collected by a digital conversational agent implemented before the PAC. In this setting, the MyRISK score was associated with the occurrence of postoperative complications at 6 months after surgery. The strength of its predictive value was increased using a machine learning model to reclassify patients classified as intermediate risk. The excellent levels of satisfaction and usability obtained from patients encourage us to develop and use this digital solution in health care. Further studies evaluating the overall use of the MyRISK score are necessary before using this digital stratification method to guide patients to teleconsultation or face-to-face consultation or to provide a perioperative personalized care pathway for patients at highest risk.

Acknowledgments

The authors would like to thank Olivier Thuillart and Jean-Louis Fraysse from BOTdesign (Toulouse, France), the company that created the digital companion *Medical Assistant Experience*. Support was provided solely from department sources. This work should be attributed to the Département d'Anesthésie Réanimation et Médecine Périopératoire, Centre Hospitalier Universitaire de Toulouse Purpan, Toulouse, France.

Authors' Contributions

FF, RL, and PF participated in the design of the study, performed all of the statistical analyses, and wrote the manuscript. ED implemented the machine learning model. MK, TG, AP, and VM participated in the design of the study and helped to draft the manuscript. AF, LB, CB, and RM performed all of the preanesthetic consultations. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Patient responding to the preanesthetic digital conversational agent.
[PNG File , 809 KB - periop_v6i1e39044_app1.png]

Multimedia Appendix 2
Example of patient completing the digital questionnaire.
[MOV File , 11775 KB - periop_v6i1e39044_app2.mov]

XSL•FO
RenderX

References

1. Recommendations concerning relations between RA and other specialties or health professionals. National Order of Doctors. 2001 Dec. URL: https://sfar.org/recommandations-concernant-les-relations-entre-ar-et-autres-specialites-ou-professionnels-de-sante/ [accessed 2022-11-29]
2. Ministère des affaires sociales, de la santé et de la ville. Décret no 94-1054 du 5 décembre 1994 relatif aux conditions techniques de fonctionnement des établissements de santé en ce qui concerne la pratique de l'anesthésie et modifiant le code de la santé publique (troisième partie : décrets) NOR : SPSH9403474D. J officiel de la République française 1994 Dec 8:17383-17385.
3. Dripps RD. Pharmacodynamics of human disease. 2. Therapeutic application of opiates. Postgrad Med 1963 Nov;34:520-524. [doi: 10.1080/00325481.1963.11694918] [Medline: 14054624]
4. Auroy Y, Benhamou D. [The pre-anesthetic consultation]. Rev Prat 2010 Nov 20;60(9):1255, 1258-1255, 1259. [Medline: 21328860]
5. Wijeysundera DN. Preoperative consultations by anesthesiologists. Curr Opin Anaesthesiol 2011 Jun;24(3):326-330. [doi: 10.1097/ACO.0b013e328345d844] [Medline: 21451401]
6. Silveira SQ, da Silva LM, Gomes RF, de Campos Vieira Abib A, Vieira JE, Ho AM, et al. An evaluation of the accuracy and self-reported confidence of clinicians in using the ASA-PS Classification System. J Clin Anesth 2022 Aug;79:110794. [doi: 10.1016/j.jclinane.2022.110794] [Medline: 35367956]
7. Hurwitz EE, Simon M, Vinta SR, Zehm CF, Shabot SM, Minhajuddin A, et al. Adding examples to the ASA-physical status classification improves correct assignment to patients. Anesthesiology 2017 Apr;126(4):614-622 [FREE Full text] [doi: 10.1097/ALN.0000000000001541] [Medline: 28212203]
8. Composition of the working group. Société française d'anesthésie réanimation. URL: https://sfar.org/dossier-anesthesique/ [accessed 2022-11-29]
9. Bibault J, Chaix B, Guillemassé A, Cousin S, Escande A, Perrin M, et al. A chatbot versus physicians to provide information for patients with breast cancer: blind, randomized controlled noninferiority trial. J Med Internet Res 2019 Nov 27;21(11):e15787 [FREE Full text] [doi: 10.2196/15787] [Medline: 31774408]
10. Gerbert B, Bronstone A, Pantilat S, McPhee S, Allerton M, Moe J. When asked, patients tell: disclosure of sensitive health-risk behaviors. Med Care 1999 Jan;37(1):104-111. [doi: 10.1097/00005650-199901000-00014] [Medline: 10413398]
11. Goodhart IM, Ibbotson V, Doane A, Roberts B, Campbell MJ, Ross RJM. Hypopituitary patients prefer a touch-screen to paper quality of life questionnaire. Growth Horm IGF Res 2005 Dec;15(6):384-387. [doi: 10.1016/j.ghir.2005.08.002] [Medline: 16183313]
12. Menachemi N, Chukmaitov A, Saunders C, Brooks RG. Hospital quality of care: does information technology matter? The relationship between information technology adoption and quality of care. Health Care Manage Rev 2008;33(1):51-59. [doi: 10.1097/01.HMR.0000304497.89684.36] [Medline: 18091444]
13. Miller ET, Neal DJ, Roberts LJ, Baer JS, Cressler SO, Metrik J, et al. Test-retest reliability of alcohol measures: is there a difference between internet-based assessment and traditional methods? Psychol Addict Behav 2002 Mar;16(1):56-63. [Medline: 11934087]
14. VanDenKerkhof EG, Goldstein DH, Blaine WC, Rimmer MJ. A comparison of paper with electronic patient-completed questionnaires in a preoperative clinic. Anesth Analg 2005 Oct;101(4):1075-1080. [doi: 10.1213/01.ane.0000168449.32159.7b] [Medline: 16192524]
15. Ferré F, Boeschlin N, Bastiani B, Castel A, Ferrier A, Bosch L, et al. Improving provision of preanesthetic information through use of the digital conversational agent "MyAnesth": prospective observational trial. J Med Internet Res 2020 Dec 04;22(12):e20455 [FREE Full text] [doi: 10.2196/20455] [Medline: 33275108]
16. Howell M, Hood AJ, Jayne DG. Use of a patient completed iPad questionnaire to improve pre-operative assessment. J Clin Monit Comput 2017 Feb;31(1):221-225. [doi: 10.1007/s10877-015-9818-0] [Medline: 26715416]
17. Molina CS, Thakore RV, Blumer A, Obremskey WT, Sethi MK. Use of the national surgical quality improvement program in orthopaedic surgery. Clin Orthop Relat Res 2015 May;473(5):1574-1581 [FREE Full text] [doi: 10.1007/s11999-014-3597-7] [Medline: 24706043]
18. Maurice-Szamburski A, Loundou A, Capdevila X, Bruder N, Auquier P. Validation of the French version of the Amsterdam preoperative anxiety and information scale (APAIS). Health Qual Life Outcomes 2013 Oct 07;11:166 [FREE Full text] [doi: 10.1186/1477-7525-11-166] [Medline: 24099176]
19. Lee TH, Marcantonio ER, Mangione CM, Thomas EJ, Polanczyk CA, Cook EF, et al. Derivation and prospective validation of a simple index for prediction of cardiac risk of major noncardiac surgery. Circulation 1999 Sep 07;100(10):1043-1049. [doi: 10.1161/01.cir.100.10.1043] [Medline: 10477528]
20. Chung F, Yegneswaran B, Liao P, Chung SA, Vairavanathan S, Islam S, et al. STOP questionnaire: a tool to screen patients for obstructive sleep apnea. Anesthesiology 2008 May;108(5):812-821 [FREE Full text] [doi: 10.1097/ALN.0b013e31816d83e4] [Medline: 18431116]
21. Niederberger M, Spranger J. Delphi technique in health sciences: a map. Front Public Health 2020;8:457 [FREE Full text] [doi: 10.3389/fpubh.2020.00457] [Medline: 33072683]

XSL•FO
RenderX

22. Story DA. Postoperative complications in elderly patients and their significance for long-term prognosis. Curr Opin Anaesthesiol 2008 Jun;21(3):375-379. [doi: 10.1097/ACO.0b013e3282f889f8] [Medline: 18458558]
23. Benumof JL. Obesity, sleep apnea, the airway and anesthesia. Curr Opin Anaesthesiol 2004 Feb;17(1):21-30. [doi: 10.1097/00001503-200402000-00005] [Medline: 17021525]
24. Lumb AB. Pre-operative respiratory optimisation: an expert review. Anaesthesia 2019 Jan;74 Suppl 1:43-48 [FREE Full text] [doi: 10.1111/anae.14508] [Medline: 30604419]
25. Chan MT, Wang CY, Seet E, Tam S, Lai HY, Chew EF, Postoperative Vascular Complications in Unrecognized Obstructive Sleep Apnea (POSA) Study Investigators. Association of unrecognized obstructive sleep apnea with postoperative cardiovascular events in patients undergoing major noncardiac surgery. JAMA 2019 May 14;321(18):1788-1798 [FREE Full text] [doi: 10.1001/jama.2019.4783] [Medline: 31087023]
26. Riedel B, Li MH, Lee CH, Ismail H, Cuthbertson BH, Wijeysundera DN, METS Study Investigators. A simplified (modified) Duke Activity Status Index (M-DASI) to characterise functional capacity: a secondary analysis of the Measurement of Exercise Tolerance before Surgery (METS) study. Br J Anaesth 2021 Jan;126(1):181-190 [FREE Full text] [doi: 10.1016/j.bja.2020.06.016] [Medline: 32690247]
27. Maddox TM. Preoperative cardiovascular evaluation for noncardiac surgery. Mt Sinai J Med 2005 May;72(3):185-192. [Medline: 15915313]
28. Blitz JD, Shoham MH, Fang Y, Narine V, Mehta N, Sharma BS, et al. Preoperative renal insufficiency: underreporting and association with readmission and major postoperative morbidity in an academic medical center. Anesth Analg 2016 Dec;123(6):1500-1515. [doi: 10.1213/ANE.0000000000001573] [Medline: 27861446]
29. Inouye SK, Westendorp RG, Saczynski JS. Delirium in elderly people. Lancet 2014 Mar;383(9920):911-922. [doi: 10.1016/s0140-6736(13)60688-1]
30. Khwaja A. KDIGO clinical practice guidelines for acute kidney injury. Nephron Clin Pract 2012;120(4):c179-c184 [FREE Full text] [doi: 10.1159/000339789] [Medline: 22890468]
31. Devereaux PJ, Sessler DI. Cardiac complications in patients undergoing major noncardiac surgery. N Engl J Med 2015 Dec 03;373(23):2258-2269. [doi: 10.1056/NEJMra1502824] [Medline: 26630144]
32. Sellers D, Srinivas C, Djaiani G. Cardiovascular complications after non-cardiac surgery. Anaesthesia 2018 Jan;73 Suppl 1:34-42 [FREE Full text] [doi: 10.1111/anae.14138] [Medline: 29313903]
33. Writing Committee for the VISION Study Investigators, Devereaux PJ, Biccard BM, Sigamani A, Xavier D, Chan MT, et al. Association of postoperative high-sensitivity troponin levels with myocardial injury and 30-day mortality among patients undergoing noncardiac surgery. JAMA 2017 Apr 25;317(16):1642-1651. [doi: 10.1001/jama.2017.4360] [Medline: 28444280]
34. Lin M, Kamel H, Singer DE, Wu Y, Lee M, Ovbiagele B. Perioperative/postoperative atrial fibrillation and risk of subsequent stroke and/or mortality. Stroke 2019 Jun;50(6):1364-1371. [doi: 10.1161/STROKEAHA.118.023921] [Medline: 31043148]
35. Bateman BT, Schumacher HC, Wang S, Shaefi S, Berman MF. Perioperative acute ischemic stroke in noncardiac and nonvascular surgery: incidence, risk factors, and outcomes. Anesthesiology 2009 Feb;110(2):231-238 [FREE Full text] [doi: 10.1097/ALN.0b013e318194b5ff] [Medline: 19194149]
36. Mashour GA, Shanks AM, Kheterpal S. Perioperative stroke and associated mortality after noncardiac, nonneurologic surgery. Anesthesiology 2011 Jun;114(6):1289-1296 [FREE Full text] [doi: 10.1097/ALN.0b013e318216e7f4] [Medline: 21478735]
37. Warner DO. Preventing postoperative pulmonary complications: the role of the anesthesiologist. Anesthesiology 2000 May;92(5):1467-1472 [FREE Full text] [doi: 10.1097/00000542-200005000-00037] [Medline: 10781293]
38. Gade IL, Kold S, Severinsen MT, Kragholm KH, Torp-Pedersen C, Kristensen SR, et al. Venous thromboembolism after lower extremity orthopedic surgery: a population-based nationwide cohort study. Res Pract Thromb Haemost 2021 Jan;5(1):148-158 [FREE Full text] [doi: 10.1002/rth2.12449] [Medline: 33537539]
39. Brooke J. SUS: a retrospective. J Usab Stud archive 2013;8:29-40 [FREE Full text]
40. Brooke J. SUS: a 'quick and dirty' usability scale. In: Usability Evaluation In Industry. Boca Raton, Florida, United States: CRC Press; 1996.
41. Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. Mach Learn 2002;46:389-422 [FREE Full text]
42. Lundberg S, Lee S. A unified approach to interpreting model predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017 Presented at: NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems; Dec 4 - 9, 2017; Long Beach California USA.
43. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016 Presented at: KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; Aug 13 - 17, 2016; San Francisco California USA. [doi: 10.1145/2939672.2939785]
44. Zuidema X, Tromp Meesters RC, Siccama I, Houweling PL. Computerized model for preoperative risk assessment. Br J Anaesth 2011 Aug;107(2):180-185 [FREE Full text] [doi: 10.1093/bja/aer151] [Medline: 21659404]

XSL•FO
RenderX

45. Enneking FK, Radhakrishnan NS, Berg K, Patel S, Wishin JM, Vasilopoulos T. Patient-centered anesthesia triage system predicts ASA physical status. Anesth Analg 2017 Jun;124(6):1957-1962. [doi: 10.1213/ANE.0000000000001712] [Medline: 28257309]
46. Teleconsultation: a practice that takes hold over time. l'assurance maladie. 2020 Sep 16. URL: https://assurance-maladie.ameli.fr/presse/2020-09-16-cp-teleconsultation-anniversaire [accessed 2022-11-29]
47. Wosik J, Fudim M, Cameron B, Gellad ZF, Cho A, Phinney D, et al. Telehealth transformation: COVID-19 and the rise of virtual care. J Am Med Inform Assoc 2020 Jun 01;27(6):957-962 [FREE Full text] [doi: 10.1093/jamia/ocaa067] [Medline: 32311034]
48. Osman T, Lew E, Lum EP, van Galen L, Dabas R, Sng BL, et al. PreAnaesThesia computerized health (PATCH) assessment: development and validation. BMC Anesthesiol 2020 Nov 14;20(1):286 [FREE Full text] [doi: 10.1186/s12871-020-01202-8] [Medline: 33189131]
49. Meirte J, Hellemans N, Anthonissen M, Denteneer L, Maertens K, Moortgat P, et al. Benefits and disadvantages of electronic patient-reported outcome measures: systematic review. JMIR Perioper Med 2020 Apr 03;3(1):e15588 [FREE Full text] [doi: 10.2196/15588] [Medline: 33393920]
50. Kruse CS, Kristof C, Jones B, Mitchell E, Martinez A. Barriers to electronic health record adoption: a systematic literature review. J Med Syst 2016 Dec;40(12):252 [FREE Full text] [doi: 10.1007/s10916-016-0628-9] [Medline: 27714560]
51. Chikuda H, Yasunaga H, Horiguchi H, Takeshita K, Sugita S, Taketomi S, et al. Impact of age and comorbidity burden on mortality and major complications in older adults undergoing orthopaedic surgery: an analysis using the Japanese diagnosis procedure combination database. BMC Musculoskelet Disord 2013 May 28;14:173 [FREE Full text] [doi: 10.1186/1471-2474-14-173] [Medline: 23711221]
52. Quality of care perceived by the patient - PROMs and PREMs indicators: overview of foreign experiences and main lessons. HAS. 2021. URL: https://www.has-sante.fr/upload/docs/application/pdf/2021-07/rapport_panorama_proms_prems_2021.pdf [accessed 2022-11-28]

Abbreviations

ASA: American Society of Anesthesiologists
AUC: area under the receiver operating characteristic curve
MAX: Medical Assistant Experience
NRS: numeric rating scale
OR: odds ratio
PAC: preanesthetic consultation
RFE: recursive feature elimination
STOP-BANG: snoring, tiredness, observed apnea, blood pressure, BMI, age, neck circumference, and gender
SUS: system usability scale

Edited by R Lee; submitted 26.04.22; peer-reviewed by H Karim, S Azadnajafabad; comments to author 01.06.22; revised version received 02.08.22; accepted 16.08.22; published 16.01.23.

Please cite as:
Ferré F, Laurent R, Furelau P, Doumard E, Ferrier A, Bosch L, Ba C, Menut R, Kurrek M, Geeraerts T, Piau A, Minville V
Perioperative Risk Assessment of Patients Using the MyRISK Digital Score Completed Before the Preanesthetic Consultation: Prospective Observational Study
JMIR Perioper Med 2023;6:e39044
URL: https://periop.jmir.org/2023/1/e39044
doi:10.2196/39044
PMID:36645704

©Fabrice Ferré, Rodolphe Laurent, Philippine Furelau, Emmanuel Doumard, Anne Ferrier, Laetitia Bosch, Cyndie Ba, Rémi Menut, Matt Kurrek, Thomas Geeraerts, Antoine Piau, Vincent Minville. Originally published in JMIR Perioperative Medicine (http://periop.jmir.org), 16.01.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Perioperative Medicine, is properly cited. The complete bibliographic information, a link to the original publication on http://periop.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Remote Home Monitoring of Continuous Vital Sign Measurements by Wearables in Patients Discharged After Colorectal Surgery: Observational Feasibility Study

Jobbe P L Leenen[1,2,3], RN, MSc; Vera Ardesch[1,4], RN, MSc; Gijsbert Patijn[1,2], MD, PhD

[1]Connected Care Center, Isala, Zwolle, Netherlands
[2]Department of Surgery, Isala, Zwolle, Netherlands
[3]Isala Academy, Isala, Zwolle, Netherlands
[4]Flexpool General Wards, Department of Care Support, Isala, Zwolle, Netherlands

Corresponding Author:
Jobbe P L Leenen, RN, MSc
Connected Care Center
Isala
Dr. van Heesweg 2
Zwolle, 8025 AB
Netherlands
Phone: 31 640275833
Email: j.p.l.leenen@isala.nl

Abstract

Background: Hospital stays after colorectal surgery are increasingly being reduced by enhanced recovery and early discharge protocols. As a result, postoperative complications may frequently manifest after discharge in the home setting, potentially leading to emergency room presentations and readmissions. Virtual care interventions after hospital discharge may capture clinical deterioration at an early stage and hold promise for the prevention of readmissions and overall better outcomes. Recent technological advances have enabled continuous vital sign monitoring by wearable wireless sensor devices. However, the potential of these devices for virtual care interventions for patients discharged after colorectal surgery is currently unknown.

Objective: We aimed to determine the feasibility of a virtual care intervention consisting of continuous vital sign monitoring with wearable wireless sensors and teleconsultations for patients discharged after colorectal surgery.

Methods: In a single-center observational cohort study, patients were monitored at home for 5 consecutive days after discharge. Daily vital sign trend assessments and telephone consultations were performed by a remote patient-monitoring department. Intervention performance was evaluated by analyzing vital sign trend assessments and telephone consultation reports. Outcomes were categorized as “no concern,” “slight concern,” or “serious concern.” Serious concern prompted contact with the surgeon on call. In addition, the quality of the vital sign data was determined, and the patient experience was evaluated.

Results: Among 21 patients who participated in this study, 104 of 105 (99%) measurements of vital sign trends were successful. Of these 104 vital sign trend assessments, 68% (n=71) did not raise any concern, 16% (n=17) were unable to be assessed because of data loss, and none led to contacting the surgeon. Of 62 of 63 (98%) successfully performed telephone consultations, 53 (86%) did not raise any concerns and only 1 resulted in contacting the surgeon. A 68% agreement was found between vital sign trend assessments and telephone consultations. Overall completeness of the 2347 hours of vital sign trend data was 46.3% (range 5%-100%). Patient satisfaction score was 8 (IQR 7-9) of 10.

Conclusions: A home monitoring intervention of patients discharged after colorectal surgery was found to be feasible, given its high performance and high patient acceptability. However, the intervention design needs further optimization before the true value of remote monitoring for early discharge protocols, prevention of readmissions, and overall patient outcomes can be adequately determined.

(JMIR Perioper Med 2023;6:e45113) doi:10.2196/45113

XSL•FO
RenderX

KEYWORDS

telemedicine; remote monitoring; home monitoring; virtual care intervention; colorectal surgery; continuous vital signs monitoring; wearable wireless monitoring; clinical deterioration; readmission; virtual care; online intervention; e-health; remote monitoring; telehealth; vital signs

Introduction

Colorectal surgery is known for high complication and readmission rates [1-5]. In the last decade, enhanced recovery programs, such as the Enhanced Recovery After Surgery (ERAS) program, have been adopted widely and have resulted in significantly shorter hospital lengths of stay, with discharge as early as postoperative day 1 or 2 [6-8]. Serious postoperative complications such as anastomotic leak, abscess, ileus, thrombosis, or surgical site infection may therefore manifest themselves in the home setting [5,9]. Follow-up of these patients is generally limited to outpatient clinic visits that do not take place until several weeks after discharge. Late recognition of signs and symptoms by patients may cause delayed detection and lead to inferior clinical outcomes and readmissions [1,5,10].

Virtual care interventions such as remote patient monitoring with mobile health apps have the potential to further reduce lengths of hospital stay and prevent unnecessary readmissions, reduce emergency department visits, and help mitigate increasing nursing staff shortages and overall health care costs [11-13]. Such interventions require only limited investment in time and money if deployed at sufficient scale [14]. Virtual care interventions may therefore help to remove barriers to further expand early discharge protocols.

Recent technological advances have enabled continuous vital sign monitoring by wearable wireless sensor devices [15,16]. These devices have been shown to be able to accurately detect deviating vital sign trends [15,17]. Such technology may allow patients to safely recover at home while being carefully monitored with the intention to capture possible clinical deterioration at an early stage.

Several studies have demonstrated the feasibility of continuous vital sign monitoring in the hospital with wireless wearables [18-21], but there are only a few studies on remote home monitoring after discharge, and these have mostly been limited to intermittent monitoring of vital signs [11,22]. Only one study showed that continuous vital sign monitoring was technically feasible and well accepted by patients discharged after esophageal surgery [23]. It is unknown if a comparable intervention is feasible in patients after colorectal surgery. Therefore, the objective of this study was to determine the feasibility of a virtual care intervention consisting of continuous vital sign monitoring with wearable wireless sensors and teleconsultations supported by a central, remote patient-monitoring department for patients discharged after colorectal surgery.

Methods

Study Design and Setting

A single-center observational cohort study was conducted in May and June 2022 in a 1250-bed teaching hospital in the Netherlands. This study is reported in accord with the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines [24].

Participants

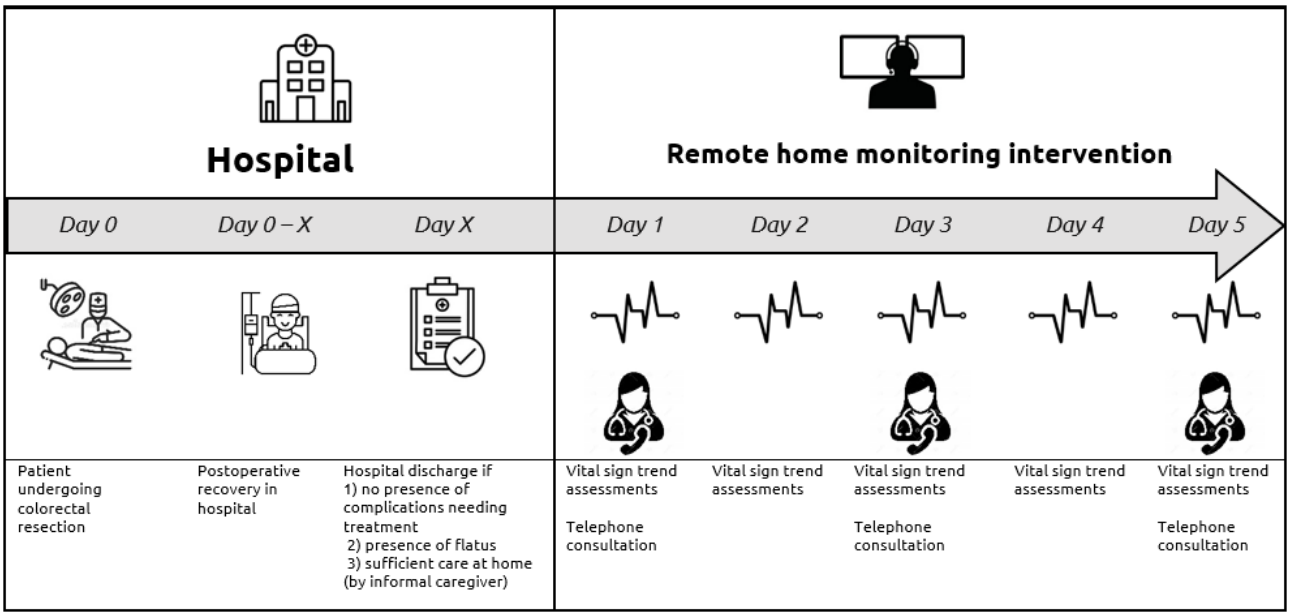
Patients scheduled for elective colorectal resection in May and June 2022 were approached for consent for the remote monitoring intervention during the preadmission call by a nurse. Inclusion criteria were as follows: age ≥18 years, elective colorectal resection, primary anastomosis, admission to the participating surgical ward, an uncomplicated clinical course (duration of admission <7 days), presence of an adequate caregiver at home, possession of a mobile phone, and no cognitive impairments. Exclusion criteria were as follows: unable to wear a continuous monitoring device due to a pacemaker or allergy, no desire for treatment (or no desire for referral) in the event of clinical deterioration, cognitive impairment at discharge, discharge to a rehabilitation or nursing home, physical limitations that would hinder participation, and insufficient command of the Dutch language.

Remote Home Monitoring Intervention

The remote home monitoring intervention was developed based on a previous study [25] that evaluated in-hospital continuous vital sign monitoring developed in cooperation with an abdominal surgeon, ward nurses, and nurses at the remote monitoring department of the hospital. The intervention consisted of 5 days of follow-up with 3 consultations by phone (every other day) and daily evaluation of vital sign trends with the wearable sensor (Figure 1).

XSL•FO
RenderX

Figure 1. Overview of the care path for included patients.



The consultations by phone were performed by the specialized cardiac care nurses at the remote monitoring department, who were dedicated to and had expertise in virtual care and remote monitoring of patients. The department is operated every day of the week from 8 AM to 11 PM. The consultation consisted of an assessment of the patient's condition regarding pain symptoms, wound condition, stool, and nausea. In addition, vital sign trends measured by the wearable sensor were assessed daily and at the time of the telephone consultation.

Vital signs of patients were measured by the Conformité Européene–marked Healthdot sensor and Intellivue Guardian Solution (IGS) software system (Philips). The wireless sensor is a previously validated patch worn on the patient's chest that continuously records heart rate (HR) in beats per minute (bpm) and respiratory rate (RR) in respirations per minute (rpm) with a battery life of 14 days (Multimedia Appendix 1). The 2 vital sign measurements are transmitted wirelessly every 5 minutes through a long range, low power Internet of Things (LoRa) connection to the IGS software. Within the IGS software, vital sign trends are visualized and, complementary to the hospital Modified Early Warning Score (MEWS) protocol, the Deel EWS (D-EWS; *Deel* is Dutch for "partial"), a partial MEWS score, was aggregated every hour to promote adequate interpretation (Multimedia Appendix 2). The scores are based on the thresholds for HR and RR when sufficient data were present. There were no active alarms generated to the nurses when scores deviated. If any acute help was needed, patients were informed to call emergency services.

Study Procedures

Before the start of the study, nurses at the remote patient monitoring department were educated by the project manager (JPLL) and a colorectal surgeon (GP) in a 1.5-hour information session about colorectal resections, the most common complications, and nursing care for these patients. Moreover, the trend assessment used in this study was informed by case studies. In a 2-month period (May-June 2022), patients scheduled for colorectal resection were approached by the ward nurse, received information about the home monitoring intervention, and were asked to provide verbal informed consent during a preoperative telephone consultation. Directly postoperatively, the sensor was attached to the patient at the ward and in-hospital continuous monitoring was performed. When the in-hospital postoperative course was uncomplicated, the patient was eligible for home monitoring in this study. At discharge, the ward nurse asked the remote monitoring department to start the monitoring.

Data Collection

Baseline characteristics were collected, including age, sex, height, weight, BMI, American Society of Anesthesiologists (ASA) score, Charlson Comorbidity Index [26], type of colorectal surgery and procedure (open or laparoscopic), indication for surgery (malignant or benign), length of hospital stay in days, MEWS at discharge, heart rate and respiratory rate at discharge, and readmission ≤30 days. The primary end point was feasibility, defined by the following 3 outcomes: intervention performance, quality of vital sign data, and patient experience.

Intervention Performance

For each patient, we assessed whether the intervention was carried out according to protocol with daily vital sign assessment and telephone consultations. Daily vital sign trend assessments were scored as 0 (no cause for concern), 1 (slight concern), and 2 (serious concern). A score of 1 resulted in a wait-and-see approach and a score of 2 resulted in contact with the surgeon on call. An X was scored if the nurse was unable to assess the vital sign trend because of insufficient vital sign data. After the telephone consultation, the same scoring was done by assessing pain (rated by the Numeric Rating Scale), wound infection (redness/pain of the wound area), and experience of nausea (yes or no) and defecation (yes or no). Additionally, discrepancies in scoring between vital sign trend assessments and telephone consultations were registered. Finally, the time spent for the intervention and for each consultation was registered in time frames of 5 minutes.

XSL•FO
RenderX

Quality of the Vital Sign Data

The quality of vital sign data was defined as the proportion of trend assessments that were impossible to perform, as well as completeness of the vital sign data, the generated D-EWS scores, heart rate and respiration measurements, and the distribution of data gaps based on their duration: <1 hour, 1 to 4 hours, 4 to 8 hours, 8 to 16 hours, and 16 to 24 hours.

Patient Experience

Patient experience was measured at the end of the remote monitoring intervention by an 11-item questionnaire. Two questions assessed the overall experience of the patient and informal caregiver (when present) with a 10-point Likert scale, 7 questions assessed specific domains with a 5-point Likert scale, and 2 questions had open answers (Multimedia Appendix 3).

Statistical Analysis

Descriptive statistics were used to evaluate patient demographics and to assess the feasibility of home monitoring. Normally distributed continuous data are presented as the mean (SD). Likewise, nonnormally distributed data are presented as the median (IQR). Normality was determined with the Kolmogorov-Smirnov test and visually by a quantile-quantile plot and histogram. Nominal data were presented as frequencies (n) and percentages (%). All data were analyzed with SPSS Statistics (version 26; IBM Corp). Formal sample size calculation was challenging given the observational feasibility study design, but a sample in the range of 20 to 25 is considered adequate for this type of study [27]. Therefore, we aimed to include 20 patients and obtain data from 100 home monitoring days.

Ethical Considerations

The Medical Ethics Review Committee of Isala waived the need for ethical approval (210414). The study was conducted in accordance with the Declaration of Helsinki. Written informed consent was obtained from each participant in the study.

Results

Study Characteristics

A total of 25 patients were screened, of whom 23 (92%) agreed to participate. After 1 patient dropped out and 1 patient had in-hospital complications, a total of 21 patients participated in the study. The mean age was 67 (SD 13) years, and 8 (38%) participants were male. The mean length of stay was 4.4 (SD 2.2) days, and hemicolectomies were mostly right sided (n=8, 38%). Three patients (14%) were readmitted to the hospital within 30 days, but there were no alterations in clinical decision-making or the clinical course based on the intervention (Multimedia Appendix 4). A full description of the participants is given in Table 1.

XSL•FO
RenderX

Table 1. Patient characteristics (n=21).

Characteristic	Value
Age (years), mean (SD)	67 (13)
Sex (male), n (%)	8 (38)
Height (cm), mean (SD)	173 (9)
Weight (kg), mean (SD)	78 (15)
BMI (kg/m^2), mean (SD)	25.9 (4.2)
American Society of Anesthesiologists class, n (%)	
1-2	15 (71)
3-4	6 (29)
Charlson Comorbidity Index, mean (SD)	3.9 (2.3)
Type of surgery, n (%)	
Right hemicolectomy	8 (38)
Sigmoid resection	6 (29)
Abdominal perianal resection	3 (14)
Wig resection colon	2 (10)
Ileocecal resection	1 (5)
Low anterior resection	1 (5)
Laparoscopic procedure, n (%)	20 (95)
Indication for surgery, n (%)	
Malignant	16 (76)
Benign	5 (24)
Length of hospital stay (days), mean (SD)	4.4 (2.2)
Modified Early Warning Score at discharge, n (%)	
0	8 (38)
1	12 (57)
2	1 (5)
Heart rate at discharge (bpm), mean (SD)	81 (13)
Respiratory rate at discharge (rpm), mean (SD)	14 (1)
Readmission <30 days, n (%)	3 (14)

Intervention Performance

Of 21 patients, 1 received only telephone consultations, because the vital sign patch was removed by accident on the ward by a ward nurse. The median time spent for the home monitoring was 40 (IQR 35-40) minutes per patient, and 49% (51/104) of the daily home monitoring checks took 5 minutes or less for the nurses to perform (Table 2). A total of 104 (99%) vital sign trend assessments were performed. The majority (71/104, 68%) of assessments did not raise concerns and 16% (17/104) could not be assessed because of excessive data loss. Nonetheless, no cases were found where the nurse was seriously concerned about a vital sign trend.

Considering the telephone consultations, 62 of 63 (98%) were performed because in one case, the patient could not be reached. The median pain score was 3 (IQR 2-4) at the first consultation and 2 (IQR 1-2) at the third consultation. In 1 of 62 consultations (1.7%) the wound assessment was painful and resulted in a score of 1 (slightly concerned) for which a wait-and-see approach was taken. In 6 of 62 consultations (10%), the patients had not yet defecated. In the majority of consultations (53/62, 85%), no concerns were raised. Only 1 of 62 consultations (1.7%) resulted in contact with the surgeon on call, due to blood loss during defecation, for which a wait-and-see policy was agreed to be followed.

Considering agreement, 42 (68%) of 62 telephone consultations were in agreement with the trend assessments, and in 8 (13%) telephone consultations, agreement was not possible because of a lack of trend data. Only once, the telephone consultation raised concern but the trend assessment did not (Figure 2).

XSL•FO
RenderX

Table 2. Intervention performance.

Variable	Value
Total time spent by nurses on assessments per patient (minutes), median (IQR)	40 (35.0-40.0)
Time taken by nurses to complete daily home monitoring checks (n=104), n (%)	
≤5 minutes	51 (49)
>5 to ≤10 minutes	47 (45)
>10 to ≤15 minutes	5 (5)
25 minutes	1 (1)
Result of vital sign assessment (n=104), n (%)	
No concern	71 (68)
Slight concern	16 (15)
Concern	0 (0)
Unable to assess	17 (16)
Result of telephone consultations (n=62), n (%)	
No concern	53 (85)
Slight concern	8 (13)
Concern	1 (2)
Numeric Rating Scale pain score (range 0-10) , median (IQR)	
Day 1	3 (2-4)
Day 3	2 (1-4)
Day 5	2 (1-2)
Wound assessments of rubor/dolor (n=21), n (%)	1 (5)
Absence of defecation (n=63), n (%)	6 (10)

Figure 2. Agreement between trend assessments and telephone consultations (n=62). Values are shown as n (%). Green indicates agreement; orange indicates that there was no agreement and a wait-and-see approach was taken; red indicates that there was no agreement. X indicates that trend data were lacking. TC: telephone consultation.

		Vital signs trends assessments scores			
		0	1	2	X
TC scores	0	39 (62.9)	7 (11.3)	0 (0.0)	7 (11.3)
	1	4 (6.5)	3 (4.8)	0 (0.0)	1 (1.6)
	2	1 (1.6)	0 (0.0)	0 (0.0)	0 (0.0)

Quality of the Vital Sign Data

Monitoring data were available for 20 patients for a total of 2347 hours of vital sign data with a median of 116.6 (IQR 115.5-118.7) hours per patient (Table 3). Data gaps of 30 minutes or less occurred in 95.2% (12,402/13,039) of measurements. Gaps between 30 and 60 minutes were detected in 2.6% (338/13,039) of measurements, and gaps between 1 and 16 hours were detected in 2.3% (299/13,039) of measurements. This resulted in an overall completeness of vital sign trends of 46.3% (range 5%-100%).

Of the 13,039 completed measurements, HR measurements contained a median of 15.6% (IQR 9.7%-35%) artifact data and RR measurements contained a median of 21.8% (IQR 9.3%-25.9%) artifact data. A total of 215 D-EWS scores were calculated based on the data, of which 40 (18.6%) were 3 or higher.

XSL•FO
RenderX

Table 3. Quality of vital sign data (total vital sign measurements=13,039).

Variable	Value
Monitoring time (hours), median (IQR)	116.6 (115.5-118.7)
Monitoring time (minutes), median (IQR)	6995 (6929-7124)
Heart rate (beats per minute), mean (SD)	74 (12)
Respiratory rate (breaths per minute), mean (SD)	17 (2)
D-EWS[a](n=1153), n (%)	
0	215 (18.6)
1	701 (60.8)
2	197 (17.1)
3 or higher	40 (3.5)
Vital sign measurements per patient, median (IQR)	519 (284-1085)
Completeness of vital sign data, median (IQR)	36.6 (20.1-77.7)
Artifacts in heart rate measurements, median (IQR)	15.6 (9.7-35.0)
Artifacts in respiratory rate measurements, median (IQR)	21.8 (9.3-25.9)
Data gaps (n=13,039), n (%)	
0-30 minutes	12,402 (95.1)
30-60 minutes	338 (2.6)
60-120 minutes	169 (1.3)
2-4 hours	91 (0.7)
4-8 hours	26 (0.2)
8-16 hours	13 (0.1)

[a]D-EWS: Deel Early Warning Score (*deel* is Dutch for "partial").

Patient Experience

Twenty patients returned the questionnaire (Table 4). The median satisfaction scores were 8 (IQR 7-9) of 10 for patients and 8 (IQR 5-8) of 10 for caregivers (n=7). The majority of patients (n=18, 90%) found the sensor comfortable to wear, although several remarks were made about experiences of discomfort during lateral sleep because of the rigidity of the sensor's material (Multimedia Appendix 5). Further, the majority of patients liked to have insight in vital sign trends (n=18, 90%) and to have telephone consultations (n=17, 85%); a majority also found contact by phone adequate (n=17, 85%). Two patients (10%) felt safer with home monitoring, whereas 7 (35%) were neutral, and 11 (55%) did not feel safer. For future home monitoring interventions, 70% of patients agreed that they would want to be discharged from the hospital at an earlier stage if there were a home monitoring intervention. Possible reasons for this, based on the remarks section of the questionnaire, were that patients experienced better recovery in the home setting overall and because they experienced sleep interruptions at night while being admitted in the hospital (Multimedia Appendix 5).

XSL•FO
RenderX

Table 4. Results of the patient survey.

Question	Median score (IQR)	Responses, n (%)		
		Disagree[a]	Neutral[b]	Agree[c]
Likert-scale items (range 1 to 10)				
In general, how did you experience the home monitoring period? (n=20)	8 (7-9)	0 (0)	2 (10)	18 (90)
How did your informal caregiver experience the home monitoring period overall? (n=7)	8 (5-8)	0 (0)	2 (29)	5 (71)
Likert-scale items (range 1 to 5; n=20)				
I found the wearable sensor comfortable.	5 (4-5)	2 (10)	1 (5)	17 (85)
I liked that healthcare professionals could see my vital signs (heart rate, breathing) on a daily basis.	5 (4.3-5)	1 (5)	1 (5)	18 (90)
I need to have insight into my vital signs measurements (heartbeat, breathing).	4 (3-5)	0 (0)	8 (40)	12 (60)
I liked the telephone contact with the healthcare professionals.	5 (4-5)	0 (0)	3 (15)	17 (85)
The telephone contacts were sufficient.	4 (4-5)	2 (10)	1 (5)	17 (85)
I felt safer with the home monitoring than if I had not had it.	4 (3.3-5)	11 (55)	7 (35)	2 (10)
If in the future you were allowed to go home a day earlier with home monitoring, would you want to?	4 (3-5)	0 (0)	6 (30)	14 (70)

[a]Represents a score of 1 to 4 on the 10-item Likert scale or 1 to 2 on the 5-item Likert scale.

[b]Represents a score of 5 on the 10-item Likert scale or 3 on the 5-item Likert scale.

[c]Represents a score of 6 to 10 on the 10-item Likert scale or 4 to 5 on the 5-item Likert scale.

Discussion

Principal Findings

In this study, we found that a remote home monitoring intervention for colorectal surgical patients consisting of vital sign trend assessments and telephone consultations performed by nurses at a remote home monitoring department was feasible. The intervention performance and patient acceptability were high, whereas the quality of vital sign data was still variable.

Intervention Performance

Our results show that 5 days of follow-up with vital sign trend assessments combined with telephone consultations could be successfully performed. The agreement between these 2 methods of patient assessment was adequate (nearly 70%). In this small cohort, the majority of assessments did not raise concerns about patient recovery, and no subsequent interventions were required, but this needs confirmation in a larger cohort. Given the limited time spent by nurses delivering the intervention, this type of postdischarge monitoring may become cost-efficient when operated at sufficient scale [14], especially if it proves to facilitate even shorter hospital stays and prevent readmissions.

To achieve this, further design optimizations are first needed. Telephone consultations may be replaced by a symptom questionnaire and self-registration in a mobile digital app by patients themselves, which promotes self-management and reduces workload for the monitoring nurse [28].

Importantly, the optimal measurement frequency for the detection of deviations in a timely manner among postdischarge surgical patients at home is still unknown [17,29]. A limited number of measurements per day rather than continuous measurements may be sufficient for the timely detection of deterioration for this patient category. Also, daily assessment of vital sign trends with the D-EWS score may not be needed for the majority of patients, as these types of early warning scores are designed to provide alerts for serious clinical deterioration, which is rare postdischarge in the home setting. Comparative studies of continuous versus intermittent vital sign measurements are needed to further investigate the optimal intervention for this patient category.

Quality of Vital Sign Data

Besides the design of the intervention, the technology of wearable wireless devices for continuous vital sign home monitoring also needs further optimization. In our study, trends were generated by performing vital sign measurement every 5 minutes. Despite the relatively low amount of measurement artifacts and data gaps, completeness of data was still low (median 37% per patient). In particular, the highly variable completeness was partly caused by data gaps longer than 4 hours, which hampered adequate trend assessment by the nurses. A possible explanation for this data loss is that the connectivity of the wireless LoRa network is currently influenced by external factors, such as indoor coverage, which can be improved by installing an amplifier for home monitoring.

Patient Experiences

Patient acceptance was very high in this study, showing that patients were content to participate in this type of care after discharge. This may be explained by the limited time investment required of patients for the intervention and the additional patient-nurse interaction provided by telephone consultations

XSL•FO
RenderX

[30]. Furthermore, the majority of patients also indicated that they would not mind being discharged earlier if they received this remote home monitoring intervention as standard care, which shows this type of home monitoring intervention could be incorporated into early discharge protocols. In addition, when these vital sign measurements are recorded and submitted by the patients themselves instead of by automated technology, patients may participate more actively in recovery management, which is associated with better outcomes [31-33] and may encourage a more sustainable, healthier life style [11,31]. On the other hand, contextual factors, such as circadian rhythm and patient mobilization, are not taken into account with single measurements made by patients themselves. Use of a wearable sensor for continuous monitoring may provide a more holistic impression of activity and patients' circadian rhythm, but the significant associated cost of the sensors must be considered [14].

Comparison With Other Work

Previous literature on this topic is scarce, but our results on intervention performance and compliance are in line with 2 previous home monitoring interventions that were comparable to ours [23,34]. Our findings are in contrast with a previous systematic review showing that remote monitoring interventions using mobile health were associated with improved surveillance, earlier detection of complications, and more timely interventions, preventing further clinical decline [11]. However, the designs of the postoperative home monitoring interventions that were reviewed were highly variable, and the interventions did not use continuous vital sign measurements with wearable devices. With regard to patient acceptability and satisfaction, previously reported comparable interventions found similarly high rates [23,34,35].

Strengths and Limitations

This is one of the first studies to examine the feasibility of a home monitoring intervention for colorectal surgery patients using an available certified wearable sensor, software, and infrastructure. However, when interpreting the findings of this study, some limitations should be considered. First, although our primary aim was to determine feasibility, the generalizability of our results is limited because of the study design and relatively small sample size. Research on adequate and timely detection of clinical deterioration with remote home monitoring interventions (potentially preventing readmissions) requires large cohorts to assess the true benefits. Second, assessment of continuous vital sign trend data specifically for perioperative patients was a novel method for the monitoring care professionals, which may have influenced the adequacy of the assessments, even though the care professionals were specialized and trained in remote patient monitoring. Third, we used a self-developed questionnaire to assess patient experiences, since a validated questionnaire suitable for this type of intervention was not available. Finally, we did not include cost as an end point in this feasibility study. An early assessment of costs may be relevant for future optimization of such interventions, especially as wearable sensor technology is still associated with significant up-front investments and device cost.

Future Directions

Although these are promising initial results, further studies on the design of a remote home monitoring intervention for this patient category are needed before the true value for early discharge protocols, prevention of readmissions, and patient outcomes can be adequately determined. First, future research is needed to explore the needs of patients and health care professionals for each care pathway regarding telephone consultations and vital sign measurements. Consideration should be given to replacing telephone consultations with indirect digital contact through a mobile app for monitoring and coaching to achieve even more efficient care. Subsequently, the expansion of vital sign measurements to include other parameters, such as body temperature and mobility, may be relevant to capture the full status of the patient. In addition, the true added value of continuous versus intermittent vital sign measurements and the optimal frequency and timing of teleconsultations need to be determined for each patient category. Notifications generated by a combination of vital sign measurements and symptoms could assist monitoring professionals in clinical decision-making. Ultimately, optimal home monitoring protocols may eventually be personalized depending on surgery type, postoperative course, and individual patient characteristics, whereby deviating vital sign trend detection will be automated by algorithms that will support the monitoring department staff.

Conclusion

We found that a remote home monitoring intervention for colorectal surgical patients with wearable wireless sensors and telephone consultations was feasible, considering the high intervention performance and high patient acceptability. However, the design of a remote home monitoring intervention for this patient category should be further optimized before its true value for early discharge protocols, prevention of readmissions, and patient outcomes can be adequately determined.

Acknowledgments

The authors would like to thank all patients and the nurses at the Medical Coordination Office for participation in the study and Philips Healthcare and Isala I&I for technical preparation. The study was funded by the hospital Innovation & Science Funds (grant INNO2016).

Data Availability

The data sets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
The Philips Healthdot wearable sensor.
[DOCX File , 64 KB - periop_v6i1e45113_app1.docx]

Multimedia Appendix 2
Thresholds of the partial Early Warning (D-EWS) scores.
[DOCX File , 14 KB - periop_v6i1e45113_app2.docx]

Multimedia Appendix 3
Patient questionnaire.
[DOCX File , 17 KB - periop_v6i1e45113_app3.docx]

Multimedia Appendix 4
Case description of readmissions.
[DOCX File , 597 KB - periop_v6i1e45113_app4.docx]

Multimedia Appendix 5
Remarks of patients at the questionnaire (translated from Dutch).
[DOCX File , 13 KB - periop_v6i1e45113_app5.docx]

References

1. Hyde LZ, Al-Mazrou AM, Kuritzkes BA, Suradkar K, Valizadeh N, Kiran RP. Readmissions after colorectal surgery: not all are equal. Int J Colorectal Dis 2018 Dec;33(12):1667-1674. [doi: 10.1007/s00384-018-3150-3] [Medline: 30167778]
2. Kim JW, Mannalithara A, Sehgal M, Mithal A, Singh G, Ladabaum U. A nationwide analysis of readmission rates after colorectal cancer surgery in the US in the era of the Affordable Care Act. Am J Surg 2020 Oct;220(4):1015-1022. [doi: 10.1016/j.amjsurg.2020.04.013] [Medline: 32362379]
3. Wick E, Shore A, Hirose K, Ibrahim A, Gearhart S, Efron J, et al. Readmission rates and cost following colorectal surgery. Dis Colon Rectum 2011 Dec;54(12):1475-1479. [doi: 10.1097/DCR.0b013e31822ff8f0] [Medline: 22067174]
4. Li LT, Mills WL, White DL, Li A, Gutierrez AM, Berger DH, et al. Causes and prevalence of unplanned readmissions after colorectal surgery: a systematic review and meta-analysis. J Am Geriatr Soc 2013 Jul;61(7):1175-1181. [doi: 10.1111/jgs.12307] [Medline: 23730901]
5. Morris M, Deierhoi R, Richman J, Altom L, Hawn M. The relationship between timing of surgical complications and hospital readmission. JAMA Surg 2014 Apr;149(4):348-354. [doi: 10.1001/jamasurg.2013.4064] [Medline: 24522747]
6. Li C, Ferri LE, Mulder DS, Ncuti A, Neville A, Lee L, et al. An enhanced recovery pathway decreases duration of stay after esophagectomy. Surgery 2012 Oct;152(4):606-14; discussion 614. [doi: 10.1016/j.surg.2012.07.021] [Medline: 22943844]
7. Jones E, Wainwright T, Foster J, Smith J, Middleton R, Francis N. A systematic review of patient reported outcomes and patient experience in enhanced recovery after orthopaedic surgery. Ann R Coll Surg Engl 2014 Mar;96(2):89-94 [FREE Full text] [doi: 10.1308/003588414X13824511649571] [Medline: 24780662]
8. Varadhan KK, Neal KR, Dejong CHC, Fearon KCH, Ljungqvist O, Lobo DN. The enhanced recovery after surgery (ERAS) pathway for patients undergoing major elective open colorectal surgery: a meta-analysis of randomized controlled trials. Clin Nutr 2010 Aug;29(4):434-440. [doi: 10.1016/j.clnu.2010.01.004] [Medline: 20116145]
9. Pak H, Maghsoudi LH, Soltanian A, Gholami F. Surgical complications in colorectal cancer patients. Ann Med Surg (Lond) 2020 Jul;55:13-18 [FREE Full text] [doi: 10.1016/j.amsu.2020.04.024] [Medline: 32435475]
10. Vascular Events in Noncardiac Surgery Patients Cohort Evaluation (VISION) Study Investigators, Spence J, LeManach Y, Chan MTV, Wang CY, Sigamani A, et al. Association between complications and death within 30 days after noncardiac surgery. CMAJ 2019 Jul 29;191(30):E830-E837 [FREE Full text] [doi: 10.1503/cmaj.190221] [Medline: 31358597]
11. Dawes A, Lin A, Varghese C, Russell M, Lin A. Mobile health technology for remote home monitoring after surgery: a meta-analysis. Br J Surg 2021 Nov 11;108(11):1304-1314. [doi: 10.1093/bjs/znab323] [Medline: 34661649]
12. Keng CJS, Goriawala A, Rashid S, Goldstein R, Schmocker S, Easson A, et al. Home to stay: An integrated monitoring system using a mobile app to support patients at home following colorectal surgery. J Patient Exp 2020 Dec;7(6):1241-1246 [FREE Full text] [doi: 10.1177/2374373520904194] [Medline: 33457571]

XSL•FO
RenderX

13. Annis T, Pleasants S, Hultman G, Lindemann E, Thompson J, Billecke S, et al. Rapid implementation of a COVID-19 remote patient monitoring program. J Am Med Inform Assoc 2020 Aug 01;27(8):1326-1330 [FREE Full text] [doi: 10.1093/jamia/ocaa097] [Medline: 32392280]
14. Peters GM, Doggen CJM, van Harten WH. Budget impact analysis of providing hospital inpatient care at home virtually, starting with two specific surgical patient groups. BMJ Open 2022 Aug 01;12(8):e051833 [FREE Full text] [doi: 10.1136/bmjopen-2021-051833] [Medline: 35914920]
15. Leenen JPL, Leerentveld C, van Dijk JD, van Westreenen HL, Schoonhoven L, Patijn GA. Current evidence for continuous vital signs monitoring by wearable wireless devices in hospitalized adults: Systematic review. J Med Internet Res 2020 Jun 17;22(6):e18636 [FREE Full text] [doi: 10.2196/18636] [Medline: 32469323]
16. Areia C, Biggs C, Santos M, Thurley N, Gerry S, Tarassenko L, et al. The impact of wearable continuous vital sign monitoring on deterioration detection and clinical outcomes in hospitalised patients: a systematic review and meta-analysis. Crit Care 2021 Sep 28;25(1):351 [FREE Full text] [doi: 10.1186/s13054-021-03766-4] [Medline: 34583742]
17. Brekke IJ, Puntervoll LH, Pedersen PB, Kellett J, Brabrand M. The value of vital sign trends in predicting and monitoring clinical deterioration: A systematic review. PLoS One 2019;14(1):e0210875 [FREE Full text] [doi: 10.1371/journal.pone.0210875] [Medline: 30645637]
18. Leenen JPL, Dijkman EM, van Dijk JD, van Westreenen HL, Kalkman C, Schoonhoven L, et al. Feasibility of continuous monitoring of vital signs in surgical patients on a general ward: an observational cohort study. BMJ Open 2021 Feb 17;11(2):e042735 [FREE Full text] [doi: 10.1136/bmjopen-2020-042735] [Medline: 33597138]
19. Leenen JPL, Rasing HJM, van Dijk JD, Kalkman CJ, Schoonhoven L, Patijn GA. Feasibility of wireless continuous monitoring of vital signs without using alarms on a general surgical ward: A mixed methods study. PLoS One 2022;17(3):e0265435 [FREE Full text] [doi: 10.1371/journal.pone.0265435] [Medline: 35286354]
20. Downey C, Chapman S, Randell R, Brown J, Jayne D. The impact of continuous versus intermittent vital signs monitoring in hospitals: A systematic review and narrative synthesis. Int J Nurs Stud 2018 Aug;84:19-27 [FREE Full text] [doi: 10.1016/j.ijnurstu.2018.04.013] [Medline: 29729558]
21. Sun L, Joshi M, Khan SN, Ashrafian H, Darzi A. Clinical impact of multi-parameter continuous non-invasive monitoring in hospital wards: a systematic review and meta-analysis. J R Soc Med 2020 Jun;113(6):217-224 [FREE Full text] [doi: 10.1177/0141076820925436] [Medline: 32521195]
22. Gillis C, Gill M, Marlett N, MacKean G, GermAnn K, Gilmour L, et al. Patients as partners in Enhanced Recovery After Surgery: A qualitative patient-led study. BMJ Open 2017 Jun 24;7(6):e017002 [FREE Full text] [doi: 10.1136/bmjopen-2017-017002] [Medline: 28647727]
23. Breteler MJM, Numan L, Ruurda JP, van Hillegersberg R, van der Horst S, Dohmen DAJ, et al. Wireless remote home monitoring of vital signs in patients discharged early after esophagectomy: Observational feasibility study. JMIR Perioper Med 2020 Dec 04;3(2):e21705 [FREE Full text] [doi: 10.2196/21705] [Medline: 33393923]
24. von Elm E, Altman D, Egger M, Pocock S, Gøtzsche PC, Vandenbroucke J, STROBE Initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. BMJ 2007 Oct 20;335(7624):806-808 [FREE Full text] [doi: 10.1136/bmj.39335.541782.AD] [Medline: 17947786]
25. Leenen JPL, Rasing HJM, Kalkman CJ, Schoonhoven L, Patijn GA. Process Evaluation of a Wireless Wearable Continuous Vital Signs Monitoring Intervention in 2 General Hospital Wards: Mixed Methods Study. JMIR Nurs 2023 May 04;6:e44061. [doi: 10.2196/44061] [Medline: 37140977]
26. Charlson ME, Pompei P, Ales KL, MacKenzie C. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 1987;40(5):373-383. [doi: 10.1016/0021-9681(87)90171-8] [Medline: 3558716]
27. Bowen DJ, Kreuter M, Spring B, Cofta-Woerpel L, Linnan L, Weiner D, et al. How we design feasibility studies. Am J Prev Med 2009 May;36(5):452-457 [FREE Full text] [doi: 10.1016/j.amepre.2009.02.002] [Medline: 19362699]
28. Huygens M, Voogdt-Pruis H, Wouters M, Meurs M, van Lettow B, Kleijweg C, et al. The uptake and use of telemonitoring in chronic care between 2014 and 2019: Nationwide survey among patients and health care professionals in the Netherlands. J Med Internet Res 2021 May 03;23(5):e24908 [FREE Full text] [doi: 10.2196/24908] [Medline: 33938808]
29. Oliver A, Powell C, Edwards D, Mason B. Observations and monitoring: routine practices on the ward. Paediatr Nurs 2010 May;22(4):28-32. [doi: 10.7748/paed2010.05.22.4.28.c7738] [Medline: 20503687]
30. Lotfi M, Zamanzadeh V, Valizadeh L, Khajehgoodari M. Assessment of nurse-patient communication and patient satisfaction from nursing care. Nurs Open 2019 Jul;6(3):1189-1196 [FREE Full text] [doi: 10.1002/nop2.316] [Medline: 31367445]
31. Skolasky RL, Mackenzie EJ, Wegener ST, Riley LH. Patient activation and adherence to physical therapy in persons undergoing spine surgery. Spine (Phila Pa 1976) 2008 Oct 01;33(21):E784-E791 [FREE Full text] [doi: 10.1097/BRS.0b013e31818027f1] [Medline: 18827683]
32. Andrawis J, Akhavan S, Chan V, Lehil M, Pong D, Bozic K. Higher preoperative patient activation associated with better patient-reported outcomes after total joint arthroplasty. Clin Orthop Relat Res 2015 Aug;473(8):2688-2697 [FREE Full text] [doi: 10.1007/s11999-015-4247-4] [Medline: 25758378]

XSL•FO
RenderX

33. Dumitra T, Ganescu O, Hu R, Fiore JF, Kaneva P, Mayo N, et al. Association between patient activation and health care utilization after thoracic and abdominal surgery. JAMA Surg 2021 Jan 01;156(1):e205002 [FREE Full text] [doi: 10.1001/jamasurg.2020.5002] [Medline: 33146682]
34. Jonker L, Plas M, de Bock GH, Buskens E, van Leeuwen BL, Lahr M. Remote home monitoring of older surgical cancer patients: Perspective on study implementation and feasibility. Ann Surg Oncol 2021 Jan;28(1):67-78 [FREE Full text] [doi: 10.1245/s10434-020-08705-1] [Medline: 32602060]
35. Levine DM, Ouchi K, Blanchfield B, Saenz A, Burke K, Paz M, et al. Hospital-level care at home for acutely ill adults: A randomized controlled trial. Ann Intern Med 2020 Jan 21;172(2):77-85. [doi: 10.7326/M19-0600] [Medline: 31842232]

Abbreviations

ASA: American Society of Anesthesiologists
D-EWS: Deel Early Warning Score
ERAS: Enhanced Recovery After Surgery
HR: heart rate
IGS: Intellivue Guardian Solution
LoRa: long range, low power Internet of Things
RR: respiratory rate

Edited by T Leung; submitted 21.12.22; peer-reviewed by M Coppens, MD Ruiz Morilla; comments to author 25.02.23; revised version received 01.03.23; accepted 31.03.23; published 05.05.23.

Please cite as:
Leenen JPL, Ardesch V, Patijn G
Remote Home Monitoring of Continuous Vital Sign Measurements by Wearables in Patients Discharged After Colorectal Surgery: Observational Feasibility Study
JMIR Perioper Med 2023;6:e45113
URL: https://periop.jmir.org/2023/1/e45113
doi:10.2196/45113
PMID:37145849

©Jobbe P L Leenen, Vera Ardesch, Gijsbert Patijn. Originally published in JMIR Perioperative Medicine (http://periop.jmir.org), 05.05.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Perioperative Medicine, is properly cited. The complete bibliographic information, a link to the original publication on http://periop.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Reducing Alcohol Use Before and After Surgery: Qualitative Study of Two Treatment Approaches

Lyndsay Chapman[1], MPH; Tom Ren[2], BS; Jake Solka[1], MS; Angela R Bazzi[3,4], MPH, PhD; Brian Borsari[5,6], PhD; Michael J Mello[7,8], MPH, MD; Anne C Fernandez[1], MA, PhD

[1]Department of Psychiatry, University of Michigan Medical School, Ann Arbor, MI, United States
[2]College of Medicine, Central Michigan University, Mount Pleasant, MI, United States
[3]Herbert Wertheim School of Public Health, University of California, San Diego, San Diego, CA, United States
[4]Department of Community Health Sciences, Boston University School of Public Health, Boston, MA, United States
[5]Mental Health Service, San Francisco VA Medical Center, San Francisco, CA, United States
[6]Department of Psychiatry and Behavioral Sciences, University of California, San Francisco, San Francisco, CA, United States
[7]Department of Emergency Medicine, Rhode Island Hospital, Providence, RI, United States
[8]Department of Health Services, Policy and Practice, Brown University School of Public Health, Providence, RI, United States

Corresponding Author:
Anne C Fernandez, MA, PhD
Department of Psychiatry
University of Michigan Medical School
North Campus Research Complex
2800 Plymouth Rd, Bldg 16
Ann Arbor, MI, 48109
United States
Phone: 1 734 763 9600
Email: acfernan@umich.edu

Abstract

Background: High-risk alcohol use is a common preventable risk factor for postoperative complications, admission to intensive care, and longer hospital stays. Short-term abstinence from alcohol use (2 to 4 weeks) prior to surgery is linked to a lower likelihood of postoperative complications.

Objective: The study aimed to explore the acceptability and feasibility of 2 brief counseling approaches to reduce alcohol use in elective surgical patients with high-risk alcohol use in the perioperative period.

Methods: A semistructured interview study was conducted with a group of "high responders" (who reduced alcohol use ≥50% postbaseline) and "low responders" (who reduced alcohol use by ≤25% postbaseline) after their completion of a pilot trial to explore the acceptability and perceived impacts on drinking behaviors of the 2 counseling interventions delivered remotely by phone or video call. Interview transcripts were analyzed using thematic analysis.

Results: In total, 19 participants (10 high responders and 9 low responders) from the parent trial took part in interviews. Three main themes were identified: (1) the intervention content was novel and impactful, (2) the choice of intervention modality enhanced participant engagement in the intervention, and (3) factors external to the interventions also influenced alcohol use.

Conclusions: The findings support the acceptability of both high- and low-intensity brief counseling approaches. Elective surgical patients are interested in receiving alcohol-focused education, and further research is needed to test the effectiveness of these interventions in reducing drinking before and after surgery.

Trial Registration: ClinicalTrials.gov NCT03929562; https://clinicaltrials.gov/ct2/show/NCT03929562

(JMIR Perioper Med 2023;6:e42532) doi:10.2196/42532

KEYWORDS

alcohol use; brief intervention; surgery; preoperative; alcohol use disorder; alcohol; substance use; substance abuse; postoperative; perioperative; counseling; surgical

XSL•FO
RenderX

Introduction

High-risk alcohol use (often defined as 2 or more drinks per day [1], or a score of 5 or more on the Alcohol Use Disorders Identification Test Consumption [AUDIT-C] [2,3] screening tool) is a common preventable risk factor for postoperative complications, admission to intensive care, and longer hospital stays [2-6]. Importantly, alcohol-associated adverse surgical outcomes are not specific to certain surgeries or subpopulations; instead, they are evident across heterogeneous patients and surgery types [3,5]. Short-term abstinence from alcohol use (2 to 4 weeks) prior to surgery is linked to a lower likelihood of postoperative complications [7-13]. Likewise, abstinence of 5 to 6 weeks after surgery is recommended to reduce one's risk of experiencing complications such as delayed wound healing, infection, and impaired cardiac function [11,14]. In addition, the majority of surgical patients receive an opioid prescription after surgery [15,16], for which concurrent alcohol use is dangerous and even lethal [17,18]. Despite these recommendations and research, elective surgical patients are rarely offered alcohol-focused assessment, education, intervention, or treatment referrals prior to surgery [6,19,20]. For example, one study found only 25% of studies of alcohol use in surgical patients used standardized alcohol assessments, and a qualitative study found surgical care providers often did not have interest in screening or discussing alcohol use with patients.

Currently, the field lacks rigorous research on brief interventions to reduce high-risk alcohol use prior to surgery. Existing preoperative interventions involve pharmacotherapy and frequent in-person visits to promote abstinence, making them most appropriate for individuals at the higher end of the alcohol use and alcohol use disorder spectrum [7,9]. A broader array of treatment options are still needed to address the full spectrum of patients drinking prior to surgery. To address this gap, we developed 2 brief counseling approaches—a low-intensity "brief advice" intervention and a higher-intensity "health coaching" intervention—to reduce preoperative alcohol use. Consistent with the different preferences and implementation needs of patients and clinical staff identified in our formative research, these 2 interventions varied in content, intensity, and modality [21].

In this paper, we describe the acceptability and feasibility of the brief advice and health coaching interventions from qualitative exit interviews with treatment "high responders" (who reduced alcohol use ≥50% postbaseline) and "low responders" (who reduced alcohol use by ≤25% postbaseline). In this way, we sought to gather contextual information and potential reasons for high versus low intervention response, including participants' perspectives on the impact of specific intervention elements, modalities, and any unanticipated influences on their alcohol use related to surgery preparation or recovery.

Methods

Study Design and Sample

This paper reports results of qualitative exit interviews conducted as part of the Alcohol Screening and Preoperative Intervention Research (ASPIRE) study, a randomized pilot trial of 2 preoperative alcohol interventions [22]. ASPIRE recruited 51 participants between July 2019 and February 2021 (with a pause from April to July 2020 due to COVID-19–related reductions in elective surgeries) from a large academic medical center in the Midwestern United States. To be eligible for the parent trial, participants had to meet the following inclusion criteria: (1) be scheduled for elective or semielective surgery in the next 35 to 120 days in select subspecialties (plastic, knee or hip arthroplasty, minimally invasive, endocrine, gynecology, urology, colorectal, or hepatobiliary); (2) receive regional or general anesthesia; (3) be aged 18 to 75 years; (4) have positive or unknown alcohol use in their social history in the electronic health record; and (5) meet criteria for high-risk alcohol use (AUDIT-C score ≥5). Selected surgical types included orthopedic, plastic, urology, gynecological, hepatobiliary, and general surgery. Individuals were excluded if they were undergoing surgeries requiring only local anesthesia, could not read or understand English, or had substantial cognitive impairment or evidence of psychotic symptoms (ie, delusions or hallucinations). Research staff contacted potential participants by phone (call or text). Eligible participants provided written informed consent for all trial procedures including qualitative exit interviews. Randomization into the brief advice or health coaching intervention conditions was stratified based on sex and alcohol screening scores.

Ethics Approval

The University of Michigan's Institutional Review Board approved all study protocols (HUM00156743). This is a registered clinical trial (ClinicalTrials.gov NCT03929562).

Study Interventions

Health Coaching

The health coaching intervention consisted of a trained, graduate-level health coach delivering 2 telehealth intervention sessions 4 and 2 weeks prior to surgery. Participants had the choice of taking part in-person (prepandemic only) or through teleconferencing compliant with the US Health Insurance Portability and Accountability Act. Sessions lasted 45 minutes each. Intervention content was based on principles of health coaching, an evidence-based, collaborative care approach that partners patients with coaches who teach strategies to self-manage health behavior. Key features include motivational interviewing, goal setting, and discussion of strategies for alcohol use reduction or cessation [23]. Our health coaching sessions also included a 4-page visual and text-based session guide (emailed to participants in advance) that included health education, personalized feedback, and tailored surgical risk messaging based on the health belief model [24,25]. Health coaching sessions introduced the concept of "prehabilitation," which involves improving health through behavior change before surgery in order to help improve postoperative outcomes.

XSL•FO
RenderX

This message framing emphasized two main health belief factors: (1) the desire to avoid a health problem (ie, surgical complications) and (2) the belief that a given behavior (ie, alcohol cessation) can prevent the problem. Coaching conversations were designed to increase motivation to reduce pre- and postoperative alcohol use. These sessions allowed participants to reflect on and discuss their alcohol use and produce personal perioperative health goals.

Brief Advice

The brief advice intervention consisted of one 10-minute phone session led by the same graduate-level health coach 4 weeks prior to surgery. The phone call was accompanied by a 2-page handout (emailed to participants prior to the session) that included information about participants' reported alcohol use from baseline surveys, educational information about alcohol and surgical health, and advice to stop alcohol use for 4 weeks prior to surgery and for 6 weeks after surgery. Treatment resources and alcohol withdrawal information were included in the handout.

Quantitative Assessment and Responder Classification

The ASPIRE pilot trial involved 3 quantitative assessment time points: at baseline and 1- and 4-months postbaseline. After the final (4-month) follow-up, treatment response and low response were assessed by calculating percent change in alcohol use at follow-up time points relative to baseline.

Alcohol Use

Changes in alcohol use were assessed at baseline and follow-up visits using the AUDIT-C [26] and a web-based version of the Timeline Follow Back (TLFB) [27]. The 3-item AUDIT-C assesses alcohol quantity and frequency with an overall score ranging from 0 to 12 (higher scores reflect higher levels of alcohol use). The TLFB uses a calendar format on which participants self-report the number of standard drinks (approximately 14 g of alcohol) consumed each day to yield measures of average drinks per week and percent of days abstinent. Our assessment time frames included the past 3 months for baseline and final follow-up visits, and past month at the 1-month follow-up visit.

Responder Classification

Those who reduced alcohol use more than ≥50% relative to baseline on both the AUDIT-C and TLFB (average drinks per week) were classified as high responders. Those who reduced alcohol use by ≤25% relative to baseline on both the AUDIT-C and TLFB were classified as low responders. Percentage cutoffs were data driven, chosen based on the distribution of change in alcohol use at follow-up. We used 2 measures to classify responder status to ensure the highest level of certainty in our responder and low-responder classification. These 2 measures are highly correlated but also provide slightly different alcohol use data for decision-making.

Qualitative Data Collection

Following completion of the final follow-up assessments at 4 months for the ASPIRE pilot trial, we invited participants to take part in qualitative exit interviews based on study condition (brief advice and health coaching) and treatment response. Our goal was to enroll approximately at least 25% of the trial sample of N=51. Of the 26 participants invited based on these criteria, 19 agreed and completed exit interviews.

A trained graduate-level interviewer conducted qualitative exit interviews via the internet (through Health Insurance Portability and Accountability Act–compliant teleconferencing) using a semistructured interview guide containing open-ended questions and detailed probes designed to explore how the interventions (or low-intervention factors) may have influenced alcohol and other substance use during the perioperative period (see Multimedia Appendix 1 for the guide). Questions also explored participants' perspectives on the acceptability (ie, whether the sessions and content were appropriate, appealing, and impactful) and feasibility (ie, whether participation was convenient, clear, and achievable), as well as how comfortable they felt sharing accurate and honest information about substance use with study staff and medical providers. Study staff regularly reviewed transcripts against audio recordings to confirm accuracy and deidentification.

Qualitative Coding and Data Analysis

We undertook thematic analysis of transcribed data that began with a collaborative codebook development process involving 4 research team members (the principal investigator, study coordinator, research assistant, and a lead qualitative investigator) [28-30]. First, key topics of interest were independently reviewed from the interview guide and several selected transcripts to develop a preliminary list of potential codes and definitions [31]. We met to discuss and refine this preliminary list, which we compiled into a draft codebook. Interviews were audio-recorded and transcribed verbatim by a professional transcription company that omitted potentially identifiable information. The draft codebook was then independently tested using another set of transcripts and the team members met again to discuss and refine the codebook. Through 3 additional rounds of this iterative codebook development and testing process, discrepancies in code application were identified and resolved; codes were merged, added, and removed; and code definitions were revised as necessary until consensus was reached on a final codebook. To validate the final codebook and assess consistency in code application, 2 core coders (the study coordinator and research assistant) double-coded 2 transcripts using MAXQDA (VERBI Software). After assessing and determining that there was a high degree of consistency, we divided up the remaining transcripts between the 2 coders, who then independently coded the remaining transcripts under the supervision of the principal and lead qualitative investigators. We continued holding weekly meetings to review coding progress, discuss emergent themes, and identify and clarify preliminary findings [32]. To delve deeper into emergent themes that were identified during an initial round of coding (described below), midway through data collection, new questions were added on COVID-19 experiences, and intervention changes that emerged in earlier interviews were suggested. For our thematic analysis for this paper, we focused on understanding intervention acceptability and feasibility and potential reasons for intervention response. The transcripts across these domains were then reviewed to

XSL•FO
RenderX

identify key themes, which are described below and exemplified using anonymized quotes.

Results

Overview

Among 19 participants, 9 (47%) were female, 9 (47%) were from the health coaching intervention condition, 10 (53%) were from the brief advice condition, 9 (47%) were considered high responders, and 10 (53%) were considered low responders (Table 1). From our analysis of interview data, we identified the following three themes: (1) the intervention content was novel and impactful, (2) the choice of intervention modality enhanced participant engagement in the intervention, and (3) factors external to the interventions also influenced alcohol use. These themes are detailed below and summarized in Table 2.

Table 1. Qualitative interview participant characteristics at baseline.

	Brief advice (n=10)	Health coaching (n=9)	Total (n=19)
Female sex, n (%)	4 (40)	5 (56)	9 (47)
Identifies as woman, n (%)	4 (40)	4 (44)	8 (42)
Age (years), mean (SD)	47.5 (9.75)	58 (14)	50 (17)
White, n (%)	10 (100)	8 (89)	18 (95)
Non-Hispanic, n (%)	10 (100)	9 (100)	19 (100)
Low responder, n (%)	5 (50)	5 (56)	10 (53)
Surgical category, n (%)			
Orthopedics	2 (20)	5 (56)	7 (37)
Plastic	4 (40)	3 (33)	7 (37)
Minimally invasive	2 (20)	N/A[a]	2 (11)
Other[b]	2 (20)	1 (11)	3 (16)
Drinks per week, mean (SD)	10.4 (12)	9.3 (14.8)	9.6 (14.2)
Average AUDIT-C[c] score, mean (SD)	5.9 (1.1)	6.1 (1.1)	6.0 (1.1)
Tobacco use, n (%)	1 (10)	5 (56)	6 (32)
Marijuana use, n (%)	2 (20)	4 (44)	6 (32)
Prescription opioid use, n (%)	1 (10)	3 (33)	4 (21)
Other drug use[d], n (%)	2 (20)	1 (11)	3 (16)

[a]N/A: not applicable.

[b]Urology, endocrine, and gynecology.

[c]AUDIT-C: Alcohol Use Disorders Identification Test–Consumption.

[d]Amphetamines and hallucinogens.

XSL•FO
RenderX

Table 2. Summary of qualitative findings by theme and responder status.

Theme	High responders (n=9)	Low responders (n=10)
Intervention content	• Perceived information as novel and personally relevant for improving their own surgical health. • Changed risk perceptions related to alcohol use and surgery. • Opportunity for longer discussions with a nonjudgmental health coach was important (health coaching only).	• Information is novel and believable but perceived as only relevant to others who have a more serious "alcohol problem." • The information about alcohol and surgical risk was not shared before past surgeries and therefore inconsistent.
Intervention modality	• Participants reported intervention was an acceptable modality and length.	• Phone modality limited personal connection, and length was too short to discuss a sensitive topic like alcohol use (brief advice only).
Other influential factors: study assessments	• Completing study assessments, which included completing a daily calendar of alcohol use increased awareness and motivation to change.	• Completing study assessments increased awareness of alcohol use and motivation but intervention did not (brief advice only).
Other influential factors: intervention language	• No thematic findings.	• The term intervention caused negative reactions and contributed to participants thinking their alcohol use was not heavy enough for study participation.
Other influential factors: surgery factors	• Surgery preparation and recovery influenced choice to reduce or stop drinking for a period of time independent of the intervention content. • More intensive surgeries, longer healing time, and continued use of opioid-based pain medications linked to longer alcohol abstinence.	• Less intensive surgery and shorter healing time linked to faster return to alcohol use.

Intervention Content Was Novel and Impactful

Overall, participants found both intervention conditions to be acceptable, of appropriate length, and impactful. Several participants stated that the intervention directly influenced their perioperative alcohol consumption: "It was very enjoyable. It was easy...there hasn't been a thing that's made me uncomfortable or that I haven't agreed with [in] the study." In general, participants reacted positively to the information presented in both intervention conditions. Participants considered the information regarding the risks of preoperative alcohol use on surgical outcomes to be novel (yet believable) and motivating. By receiving this information, participants described feeling more prepared for surgery and "in control" of their surgical outcomes. As one health coaching intervention high responder explained:

> *I really did appreciate having that opportunity to say, "This is something that's a big deal for my body." Anything I can do to ensure the success of the surgery is something I want to do. Then, I'm talking about [the study] with a lot of people, saying, "I never realized this, but if you are going through something medically, it's a really good idea to get yourself as ready as possible."* [55 years old, female]

Several participants commented on what they learned about the connections between alcohol use and the body's ability to heal, including alcohol's impact on the immune system. In considering what "stood out" from the brief advice condition, one high responder explained:

> *I've never had any sort of information beyond just, you know, the day of the surgery, "Don't eat anything before or after midnight." You know, the general pre-surgical conditions to get through the day. But this [intervention] made an attempt to look out for my bigger picture health for a longer period of time before and after the surgery.* [47 years old, male]

Other high responders generally perceived the information shared as part of the intervention as impactful, identifying it as one of the main reasons they reduced their preoperative alcohol use. One brief advice high responder (48 years old, male) stated "If I was not a part of the study or aware of the study, I probably would not have reduced [my drinking]." Health coaching participants described how conversations with a nonjudgmental health coach provided them an opportunity to carefully consider changing their alcohol use.

Both interventions presented information about alcohol use in the preoperative period that some participants found helped them make a "connection" between alcohol use and health, ultimately changing their personal risk perceptions. Participants found the concept of prehabilitation, or optimizing health *prior to* surgery, to be particularly impactful because the health coach encouraged them to view themselves as important members of their surgical teams who could enhance their own health, as opposed to passively receiving surgery. As such, they felt an improved sense of control over their surgical outcomes. As one health coaching high responder explained:

> *Participation gave me more control over the anxiety having to do with the surgery because I'm a person who likes to control things, and it just gave me a sense of control over one more aspect of the surgery that I could do the best I could do to have success.* [55 years old, female]

XSL•FO
RenderX

Among low responders in both intervention conditions, there was a sense that the information presented was true but did not apply to them personally. This was particularly true for participants on the lower end of the AUDIT-C eligibility assessment, who did not view their drinking habits as risky. These individuals felt they had been included in the study erroneously, with one health coaching low responder (71 years old, female) saying, "it just didn't feel like a good fit for me." A brief advice low responder (47 years old, male) further elaborated that he thought the content was "focused on people that drink a lot all the time...while I don't have a problem having a beer or two on the weekends, it's really not an issue for me."

Some participants who had undergone major surgeries in the past described never receiving information about alcohol use and related surgical risks. Accordingly, some participants mentioned feeling surprised that their doctors had not communicated this kind of information to them in the past. One brief advice low responder (22 years old, male) explained that he would have appreciated receiving better information from his surgical team given the significance of the topic, saying, "I've had two surgeries in the past on my knee...these issues literally never came up. I was never briefed on what complications might arise around alcohol or tobacco use, so that was new."

The Choice of Intervention Modality Enhanced Participant Engagement in the Intervention

By design, intervention modality differed between conditions. While the brief advice condition was only delivered by phone, health coaching participants were given a choice between phone or face-to-face sessions (by videoconferencing or, prior to COVID-19, in-person). Overall, it appeared that having some choice in intervention modality increased participants' satisfaction; while nearly all health coaching participants were satisfied with their modality of choice, brief advice participants, who had no choice, indicated mixed experiences and dissatisfaction with phone participation.

Some health coaching participants who chose face-to-face intervention delivery expressed high satisfaction with this modality, explaining that they felt more engaged with the information presented, took intervention recommendations more seriously, and were more open and honest because it felt more "personal." As one health coaching high responder (58 years old, male) explained, "You're talking to somebody, a human, and it just makes you want to participate and follow the rules, so [there's] more accountability." Most face-to-face sessions involved videoconferencing, which participants could schedule around their work and personal lives; several noted that this flexibility enabled them to save time and energy compared with traveling to in-person sessions. They were also able to participate in a setting of their choice, which in addition to being convenient, was more comfortable. Instances of dissatisfaction included one participant stating the sessions were "a bit too long" and another participant who would have preferred in-person sessions, but this was impossible due to COVID-19 pandemic restrictions at that time.

In contrast, brief advice participants' experiences with phone sessions were more mixed, possibly because they lacked any choice in intervention modality. While some participants discussed specific advantages of phone sessions (eg, greater ease of use and convenience over videoconferencing or traveling to an in-person appointment), others felt they would have been more comfortable face-to-face, with one participant (48 years old, male) saying he would be "a lot more engaged [having] a video in front of me and [knowing] I'm being watched at the same time." Furthermore, they perceived a poorer connection between themselves and the interventionist, viewing the phone modality as "un-intimate." As a brief advice low responder explained:

> *[I was] probably not as comfortable, for whatever reason. It was a phone conversation. I was in a parking lot during work, in between meetings, in my car. I guess the setting wasn't super great, and now I'm at home [during qualitative interview]. I'm sitting down, I'm not on my phone. Those are all factors.* [48 years old, female]

Most felt that the length of health coaching sessions (~45 minutes) was appropriate and afforded them sufficient time for discussion without feeling rushed while also showing respect for their time. In contrast, while most participants in the brief advice condition felt that the session length was acceptable, some felt the length was too short to feel comfortable opening up about the sensitive topic of alcohol use.

Factors External to the Interventions Also Influenced Alcohol Use

There were several factors external to the intervention content that influenced participants' alcohol consumption, including (1) self-report and tracking of alcohol use for the ASPIRE trial assessments, (2) "intervention" language used in the study, and (3) surgical experiences.

Self-Report and Tracking

Assessments for the ASPIRE study included self-reported alcohol consumption using a 3-month retrospective TLFB calendar procedure, which affected participants in 2 ways. First, high responders described that knowing the study team would review their data and goals promoted a sense of "accountability" and not wanting to be caught "breaking the rules" by drinking at higher levels. One health coaching participant (63 years old, male) explained "I think setting the goals and following through, knowing that someone was gonna get it at the end of the time period.... That helped a lot." Second, self-report and tracking gave some participants an improved awareness of their drinking habits by enabling them to visually recognize behaviors they wanted to change. Some discussed how the calendar helped reveal patterns they had not noticed before, as one brief advice low responder (43 years old, female) explained: "When you look at that calendar, after you fill it out, and go, oh my God, I drank every day for two weeks...there's no need to be doing it every day. It just seems excessive." This was especially true for those in the brief advice condition, as study assessment activities, which totaled ~2.5 hours across the duration of the ASPIRE study, were proportionally much longer than the ~10-minute intervention session. As one brief advice high responder (43 years old, female) explained, "I don't think the

XSL•FO
RenderX

intervention content was that impactful. I think it was more of just the [activities], doing the survey, kind of as a journal, that had more of an impact on me."

"Intervention" Language Used in the Study

Some participants reacted negatively to language used during recruitment and study implementation. Study materials initially referred to both conditions as "interventions," which implied to participants that they had an alcohol-related "problem" significant enough to require help. As one health coaching low responder (71 years old, female) stated, "I mean, it's hard not to feel defensive when somebody says, 'Oh, well, you're being put in an intervention.'" For some participants, this term carried negative connotations leading them to question whether they were a "good fit" for the study, particularly if they did not see themselves as having a "problem." As one health coaching low responder (71 years old, female) explained:

> *I was a little taken aback by this sense of, oh my God, you know, an "intervention." Like, where did that come from, you know? This is not something that's ever been a particular problem or concern, so it was just a little bit of a surprise.... And all of a sudden, I'm in this category like, "You need an intervention." I was like, "What? Excuse me? Hello? No, don't think so." But anyway, I was mildly taken aback by all that.*

Surgical Experiences

Most study participants viewed their surgeries as an important life event; some further recognized surgery as dangerous and requiring careful preparation and viewed alcohol consumption as harmful for surgical recovery even before the intervention. For these individuals, reducing alcohol use prior to surgery seemed intuitive, and they believed they would have done so even without the intervention. One brief advice high responder explained:

> *I know [surgery is] dangerous, just baseline dangerous, so the better health you can be in, the better the surgery will go and the recovery. I guess I was already at that point [of wanting to be healthier for the surgery], so I was—I don't think [the intervention] had any impact.* [47 years old, male]

Postoperatively, recovery prevented many participants' alcohol consumption for at least some period due to not feeling well enough to eat, drink, or ambulate. As one brief advice high responder explained:

> *Half the time, I didn't feel like even eatin' or doin' anything, so that might not be—it might not be the study that made me do what I did. It might have been the pain level and other factors.* [47 years old, male]

Others intuitively recognized that drinking could negatively impact their recovery and believed they would have reduced their alcohol consumption even without the intervention. Some participants also decided to abstain from alcohol beyond 6 weeks after surgery to continue improving their health. Others recognized that mixing alcohol and opioid-based medications is dangerous and chose to abstain from alcohol in the postoperative period for such reasons. Those with easier, shorter recovery periods did not view surgery as having a large impact on their decision to drink alcohol one way or another. Low responders began to drink alcohol as soon as they felt better from surgery, even within 6 weeks, as one brief advice participant explained:

> *[Alcohol] was non-existent the first few days after [the surgery], and then it picked back up shortly after...once I was off [OxyContin], then I was drinking.* [43 years old, female]

Discussion

Principal Findings

This study describes the experiences of participants in a preoperative pilot trial of alcohol-focused interventions of varying intensities and modalities. Participants in our qualitative exit interviews were selected to represent high responders and low responders to help better understand potential effective and ineffective elements of the interventions to inform a future sequential multiple assignment randomized controlled trial, assigning participants to various treatment options based on initial treatment response. Overall, high responders in our pilot trial described intervention content as novel and relevant. The health coaching condition, which involved motivational interviewing and goal setting, provided an opportunity to reflect on alcohol use in an open, nonjudgmental conversation with supporting materials that framed the patients' roles as active, thereby empowering them to take actions to improve their surgical outcomes. Surgery preparation and recovery appeared to motivate change in alcohol use for participants, with the intervention content helping some participants link their alcohol use to surgical health for the first time. While most participants found both interventions to be novel and motivating, a few believed the link between alcohol use and surgical health was intuitive and that they would have stopped or reduced alcohol use before and after surgery regardless of the intervention. Likewise, longer surgical recovery and prolonged medication use led to longer alcohol cessation for some participants.

Though all participants met "risky" alcohol use eligibility requirements for the study using a validated alcohol use screening and risk identification tool, many did not perceive themselves as drinking enough to put them at risk, especially in the low-responder group. Instead, some of these participants described dismissing the intervention content about alcohol use and surgical risk because they felt it did not apply to them. A modified approach that uses different terminology or alternative discussion points could help address this barrier to change. This could include acknowledging that some participants may be surprised by feedback suggesting their level of alcohol use could be linked to surgical risks and discussing their reactions to this information.

There were also aspects of study participation and surgery that influenced participants' alcohol use separate from the intervention content. There was evidence of assessment reactivity, particularly for participants in the brief advice condition for whom the baseline and follow-up study assessments represented a much larger proportion of their participation time than the actual intervention itself. Being able

XSL•FO
RenderX

to visually see patterns of alcohol use on our assessment calendar was illuminating and motivating for some, which is consistent with research finding that the simple act of measuring behaviors and comparing them to external standards or goals can result in lasting behavior change [33], with larger effects on goal attainment when outcomes are physically tracked and recorded [34].

Importantly, several participants reacted negatively to the term *intervention*, which was used to describe study treatment conditions in several study-related documents and contact scripts. For these individuals, the term *intervention* triggered negative associations with confrontational Johnson-style interventions [35] during which a person with an addiction is confronted by family and friends and sent to long-term rehabilitation, sometimes against their will. After learning of this unintended association, and the negative reaction across several participants, in July 2020, we removed this term from study documents. For example, in the consent form, we changed references from study *interventions* to study *conditions* or *treatments*. In addition, using prehabilitation program language could also reduce stigma by indicating the program's goals are to optimize a patient's health for surgery and subsequent recovery. In our next study phase, we will engage community members in reviewing all our study documents prior to recruitment for clarity and interpretation.

Limitations and Future Directions

This study has several limitations. Recruitment was paused due to the COVID-19 pandemic and the resulting cancellation of elective surgeries. When surgeries and recruitment resumed at a reduced capacity, it is possible that surgeries and participants were different in significant ways from those scheduled prior to the pandemic. Interviews were conducted by a researcher on the study team, which may have increased the risk of social desirability bias; however, it was a team member who had not been in contact with participants prior to the interview (ie, not the health coach or recruiter), and we reminded participants that the purpose of the interview was to inform intervention improvement, and they were encouraged to provide honest and critical feedback. Finally, while qualitative data can help deepen understanding of participants' intervention experience, results from this study should not be considered generalizable to broader populations. The next step in research includes a fully powered sequential multiple assignment randomized trial that will randomly assign participants to treatment at baseline and then reassign them to different treatments and or treatment intensities based on initial treatment response and low response.

Conclusions

In conclusion, through in-depth interviews with a select sample of participants in the ASPIRE pilot trial, we found that both health coaching and brief advice conditions were acceptable pending some minor improvements and modifications. Further research is needed to test the effectiveness of these interventions in reducing drinking before and after surgery, both in the short and long term. Future research and programmatic work on these topics should carefully consider the use of lay versus technical (ie, psychiatric) terminology (eg, *intervention*) to avoid unintended negative reactions. Future research should also investigate whether assessments and feedback alone, without being coupled with brief interventions or health coaching, may be sufficient in changing participant alcohol use behaviors before or after surgery.

Acknowledgments

This study was funded by a grant from the National Institute on Alcohol Abuse and Alcoholism (K23AA023869) to ACF as the principal investigator. ARB's effort was supported by a career development award from the National Institute on Drug Abuse (K01DA043412).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Patient qualitative interview guide.
[DOCX File , 56 KB - periop_v6i1e42532_app1.docx]

References

1. US Office of Disease Prevention and Health Promotion. Appendix 9. Alcohol—2015-2020 dietary guidelines | health.gov. Dietary Guidelines 2015-2020. URL: https://health.gov/our-work/food-nutrition/2015-2020-dietary-guidelines/guidelines/appendix-9/ [accessed 2020-08-09]
2. Rubinsky AD, Bishop MJ, Maynard C, Henderson WG, Hawn MT, Harris AHS, et al. Postoperative risks associated with alcohol screening depend on documented drinking at the time of surgery. Drug Alcohol Depend 2013;132(3):521-527 [FREE Full text] [doi: 10.1016/j.drugalcdep.2013.03.022] [Medline: 23683792]
3. Bradley KA, Rubinsky AD, Sun H, Bryson CL, Bishop MJ, Blough DK, et al. Alcohol screening and risk of postoperative complications in male VA patients undergoing major non-cardiac surgery. J Gen Intern Med 2011;26(2):162-169 [FREE Full text] [doi: 10.1007/s11606-010-1475-x] [Medline: 20878363]

XSL•FO
RenderX

4. Harris AHS, Reeder R, Ellerbe L, Bradley KA, Rubinsky AD, Giori NJ. Preoperative alcohol screening scores: association with complications in men undergoing total joint arthroplasty. J Bone Joint Surg Am 2011;93(4):321-327. [doi: 10.2106/JBJS.I.01560] [Medline: 21325583]
5. Eliasen M, Grønkjær M, Skov-Ettrup LS, Mikkelsen SS, Becker U, Tolstrup JS, et al. Preoperative alcohol consumption and postoperative complications: a systematic review and meta-analysis. Ann Surg 2013;258(6):930-942. [doi: 10.1097/SLA.0b013e3182988d59] [Medline: 23732268]
6. Harris AHS, Frey MS, DeBenedetti AF, Bradley KA. Alcohol misuse prevalence and associations with post-operative complications in US surgical patients: A review. Open J Surg 2008;2(1):50-58 [FREE Full text] [doi: 10.2174/1874300500802010050]
7. Fernandez AC, Claborn KR, Borsari B. A systematic review of behavioural interventions to reduce preoperative alcohol use. Drug Alcohol Rev 2015;34(5):508-520 [FREE Full text] [doi: 10.1111/dar.12285] [Medline: 26120973]
8. Egholm JW, Pedersen B, Møller AM, Adami J, Juhl CB, Tønnesen H. Perioperative alcohol cessation intervention for postoperative complications. Cochrane Database Syst Rev 2018;11(11):CD008343 [FREE Full text] [doi: 10.1002/14651858.CD008343.pub3] [Medline: 30408162]
9. Tønnesen H, Egholm JW, Oppedal K, Lauritzen JB, Madsen BL, Pedersen B. Patient education for alcohol cessation intervention at the time of acute fracture surgery: study protocol for a randomised clinical multi-centre trial on a gold standard programme (Scand-Ankle). BMC Surg 2015;15:52 [FREE Full text] [doi: 10.1186/s12893-015-0035-z] [Medline: 25925742]
10. Jayaraj R, Thomas M, Kavanagh D, d'Abbs P, Mayo L, Thomson V, et al. Study protocol: screening and treatment of alcohol-related trauma (START)—a randomised controlled trial. BMC Health Serv Res 2012;12:371 [FREE Full text] [doi: 10.1186/1472-6963-12-371] [Medline: 23106916]
11. Lauridsen SV, Thomsen T, Thind P, Tønnesen H. STOP smoking and alcohol drinking before OPeration for bladder cancer (the STOP-OP study), perioperative smoking and alcohol cessation intervention in relation to radical cystectomy: study protocol for a randomised controlled trial. Trials 2017;18(1):329 [FREE Full text] [doi: 10.1186/s13063-017-2065-6] [Medline: 28716147]
12. Snowden C, Lynch E, Avery L, Gerrand C, Gilvarry E, Goudie N, et al. Preoperative behavioural intervention versus standard care to reduce drinking before elective orthopaedic surgery (PRE-OP BIRDS): protocol for a multicentre pilot randomised controlled trial. Pilot Feasibility Stud 2018;4:140 [FREE Full text] [doi: 10.1186/s40814-018-0330-4] [Medline: 30128165]
13. Oppedal K, Møller AM, Pedersen B, Tønnesen H. Preoperative alcohol cessation prior to elective surgery. Cochrane Database Syst Rev 2012(7):CD008343. [doi: 10.1002/14651858.CD008343.pub2] [Medline: 22786514]
14. Tønnesen H. Alcohol abuse and postoperative morbidity. Dan Med Bull 2003;50(2):139-160. [Medline: 12812138]
15. Wunsch H, Wijeysundera DN, Passarella MA, Neuman MD. Opioids prescribed after low-risk surgical procedures in the United States, 2004-2012. JAMA 2016;315(15):1654-1657 [FREE Full text] [doi: 10.1001/jama.2016.0130] [Medline: 26978756]
16. Ladha KS, Neuman MD, Broms G, Bethell J, Bateman BT, Wijeysundera DN, et al. Opioid prescribing after surgery in the United States, Canada, and Sweden. JAMA Netw Open 2019;2(9):e1910734 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.10734] [Medline: 31483475]
17. Witkiewitz K, Vowles KE. Alcohol and opioid use, co-use, and chronic pain in the context of the opioid epidemic: a critical review. Alcohol Clin Exp Res 2018;42(3):478-488 [FREE Full text] [doi: 10.1111/acer.13594] [Medline: 29314075]
18. White JM, Irvine RJ. Mechanisms of fatal opioid overdose. Addiction 1999;94(7):961-972. [Medline: 10707430]
19. Aarvold A, Crofts T. Missed opportunities? Management of patients with alcohol problems in a surgical ward. Health Bull (Edinb) 2002;60(1):55-61. [Medline: 12664770]
20. Tønnesen H, Faurschou P, Ralov H, Mølgaard-Nielsen D, Thomas G, Backer V. Risk reduction before surgery. The role of the primary care provider in preoperative smoking and alcohol cessation. BMC Health Serv Res 2010;10:121 [FREE Full text] [doi: 10.1186/1472-6963-10-121] [Medline: 20462417]
21. Fernandez AC, Guetterman TC, Borsari B, Mello MJ, Mellinger J, Tonnesen H, et al. Gaps in alcohol screening and intervention practices in surgical healthcare: a qualitative study. J Addict Med 2021;15(2):113-119 [FREE Full text] [doi: 10.1097/ADM.0000000000000706] [Medline: 32769774]
22. Fernandez AC, Chapman L, Ren TY, Baxley C, Hallway AK, Tang MJ, et al. Preoperative alcohol interventions for elective surgical patients: results from a randomized pilot trial. Surgery 2022;172(6):1673-1681. [doi: 10.1016/j.surg.2022.09.012] [Medline: 36283843]
23. Olsen JM, Nesbitt BJ. Health coaching to improve healthy lifestyle behaviors: an integrative review. Am J Health Promot 2010;25(1):e1-e12. [doi: 10.4278/ajhp.090313-LIT-101] [Medline: 20809820]
24. Rosenstock IM. The health belief model: explaining health behavior through expectancies. In: Glanz K, Rimer BK, Viswanath K, editors. Health Behavior and Health Education: Theory, Research, and Practice. The Jossey-Bass Health Series. Long Beach, CA: Jossey-Bass; 1990:39-62.
25. Janz NK, Becker MH. The health belief model: a decade later. Health Educ Q 1984;11(1):1-47. [doi: 10.1177/109019818401100101] [Medline: 6392204]

XSL•FO
RenderX

26. Bush K, Kivlahan DR, McDonell MB, Fihn SD, Bradley KA. The AUDIT alcohol consumption questions (AUDIT-C): an effective brief screening test for problem drinking. Arch Intern Med 1998;158(16):1789-1795 [FREE Full text] [doi: 10.1001/archinte.158.16.1789] [Medline: 9738608]
27. Sobell LC, Sobell MB. Timeline follow-back. In: Litten RZ, Allen JP, editors. Measuring Alcohol Consumption: Psychosocial And Biochemical Methods. Totowa, NJ: Humana Press; 1992:41-72.
28. Johnson BD, Dunlap E, Benoit E. Organizing "mountains of words" for data analysis, both qualitative and quantitative. Subst Use Misuse 2010;45(5):648-670 [FREE Full text] [doi: 10.3109/10826081003594757] [Medline: 20222777]
29. MacQueen KM, McLellan E, Kay K, Milstein B. Codebook development for team-based qualitative analysis. CAM J 2016;10(2):31-36. [doi: 10.1177/1525822X980100020301]
30. DeCuir-Gunby JT, Marshall PL, McCulloch AW. Developing and using a codebook for the analysis of interview data: an example from a professional development research project. Field Methods 2011;23(2):136-155. [doi: 10.1177/1525822X10388468]
31. McLellan E, MacQueen KM, Neidig JL. Beyond the qualitative interview: data preparation and transcription. Field Methods 2003;15(1):63-84. [doi: 10.1177/1525822x02239573]
32. Ryan GW, Bernard HR. Techniques to identify themes. Field Methods 2016;15(1):85-109. [doi: 10.1177/1525822x02239569]
33. Gass JC, Funderburk JS, Shepardson R, Kosiba JD, Rodriguez L, Maisto SA. The use and impact of self-monitoring on substance use outcomes: a descriptive systematic review. Subst Abus 2021;42(4):512-526 [FREE Full text] [doi: 10.1080/08897077.2021.1874595] [Medline: 33617740]
34. Harkin B, Webb TL, Chang BPI, Prestwich A, Conner M, Kellar I, et al. Does monitoring goal progress promote goal attainment? A meta-analysis of the experimental evidence. Psychol Bull 2016;142(2):198-229 [FREE Full text] [doi: 10.1037/bul0000025] [Medline: 26479070]
35. Fernandez AC, Begley EA, Marlatt GA. Family and peer interventions for adults: past approaches and future directions. Psychol Addict Behav 2006;20(2):207-213. [doi: 10.1037/0893-164X.20.2.207] [Medline: 16784367]

Abbreviations

ASPIRE: Alcohol Screening and Preoperative Intervention Research
AUDIT-C: Alcohol Use Disorders Identification Test Consumption
TLFB: Timeline Follow Back

Edited by T Leung; submitted 14.09.22; peer-reviewed by G Mody, J Braciszewski, S Mehdipour, T Regan; comments to author 24.03.23; revised version received 14.04.23; accepted 31.05.23; published 26.07.23.

Please cite as:
Chapman L, Ren T, Solka J, Bazzi AR, Borsari B, Mello MJ, Fernandez AC
Reducing Alcohol Use Before and After Surgery: Qualitative Study of Two Treatment Approaches
JMIR Perioper Med 2023;6:e42532
URL: https://periop.jmir.org/2023/1/e42532
doi:10.2196/42532
PMID:37494103

©Lyndsay Chapman, Tom Ren, Jake Solka, Angela R Bazzi, Brian Borsari, Michael J Mello, Anne C Fernandez. Originally published in JMIR Perioperative Medicine (http://periop.jmir.org), 26.07.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Perioperative Medicine, is properly cited. The complete bibliographic information, a link to the original publication on http://periop.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Dashboard of Short-Term Postoperative Patient Outcomes for Anesthesiologists: Development and Preliminary Evaluation

Rama Syamala Sreepada[1,2], PhD; Ai Ching Chang[1,2], MEng; Nicholas C West[2], MSc; Jonath Sujan[2], MEng; Brendan Lai[2], BASc; Andrew K Poznikoff[2,3], BSc; Rebecca Munk[4], MBBS; Norbert R Froese[1,2,3], MD; James C Chen[1,3], MD; Matthias Görges[1,2], PhD

[1]Department of Anesthesiology, Pharmacology and Therapeutics, The University of British Columbia, Vancouver, BC, Canada
[2]Research Institute, BC Children's Hospital, Vancouver, BC, Canada
[3]Department of Anesthesia, BC Children's Hospital, Vancouver, BC, Canada
[4]Department of Anesthesiology, Kelowna General Hospital, Kelowna, BC, Canada

Corresponding Author:
Matthias Görges, PhD
Research Institute
BC Children's Hospital
Rm V3-317 - 950 West 28th Avenue
Vancouver, BC, V5Z 4H4
Canada
Phone: 1 6048752000 ext 5616
Email: mgoerges@bcchr.ca

Abstract

Background: Anesthesiologists require an understanding of their patients' outcomes to evaluate their performance and improve their practice. Traditionally, anesthesiologists had limited information about their surgical outpatients' outcomes due to minimal contact post discharge. Leveraging digital health innovations for analyzing personal and population outcomes may improve perioperative care. BC Children's Hospital's postoperative follow-up registry for outpatient surgeries collects short-term outcomes such as pain, nausea, and vomiting. Yet, these data were previously not available to anesthesiologists.

Objective: This quality improvement study aimed to visualize postoperative outcome data to allow anesthesiologists to reflect on their care and compare their performance with their peers.

Methods: The postoperative follow-up registry contains nurse-reported postoperative outcomes, including opioid and antiemetic administration in the postanesthetic care unit (PACU), and family-reported outcomes, including pain, nausea, and vomiting, within 24 hours post discharge. Dashboards were iteratively co-designed with 5 anesthesiologists, and a department-wide usability survey gathered anesthesiologists' feedback on the dashboards, allowing further design improvements. A final dashboard version has been deployed, with data updated weekly.

Results: The dashboard contains three sections: (1) 24-hour outcomes, (2) PACU outcomes, and (3) a practice profile containing individual anesthesiologist's case mix, grouped by age groups, sex, and surgical service. At the time of evaluation, the dashboard included 24-hour data from 7877 cases collected from September 2020 to February 2023 and PACU data from 8716 cases collected from April 2021 to February 2023. The co-design process and usability evaluation indicated that anesthesiologists preferred simpler designs for data summaries but also required the ability to explore details of specific outcomes and cases if needed. Anesthesiologists considered security and confidentiality to be key features of the design and most deemed the dashboard information useful and potentially beneficial for their practice.

Conclusions: We designed and deployed a dynamic, personalized dashboard for anesthesiologists to review their outpatients' short-term postoperative outcomes. This dashboard facilitates personal reflection on individual practice in the context of peer and departmental performance and, hence, the opportunity to evaluate iterative practice changes. Further work is required to establish their effect on improving individual and department performance and patient outcomes.

(JMIR Perioper Med 2023;6:e47398) doi:10.2196/47398

XSL•FO
RenderX

KEYWORDS

quality improvement; feedback; anesthesiologists; patient reported outcome measures; data display; user-centered design; surgical outcome; discharge; anesthesiology; postoperative care; registry; dashboard; interactive; practice; performance; patient outcome; mobile phone

Introduction

Anesthesiologists benefit from receiving feedback on their patients' outcomes and can use it to evaluate and improve their practice. The perioperative period is a data-rich environment with the potential for innovation through digital health tools and predictive analytics. Data-driven performance feedback can improve perioperative practice and outcomes [1,2], including antibiotic administration [3,4], drug costs [5], operating room booking efficiency [6], and temperature monitoring compliance [7]. Feedback is most effective when it is locally relevant and derived from a credible source [8]. Personalized feedback, provided in (near) real time [3,9], is more effective than retrospective and department-wide feedback [8]. However, evaluating postoperative care metrics can be challenging, particularly for day-case procedures: access to this information is often only available from fragmented data sources, data are difficult to access, or are presented in user-unfriendly formats.

In anesthesia, practitioners commonly work in isolation (1 anesthesiologist per patient) and do not often have a chance to compare variations in individual practitioners' anesthetic techniques and outcomes unless required by a critical or near-miss event. While providing care, anesthesiologists make multiple decisions influencing pain and nausea outcomes [10]. Longitudinal follow-up of patients is often limited to inpatients; in high turnover pediatric operating rooms, an anesthesiologist may not have time to check in on a recovering patient before they are discharged home, and postanesthetic care unit (PACU) nursing does not routinely inform an anesthesiologist when administering ordered doses of analgesic or antiemetic rescue medications. Hence, an anesthesiologist may not know how their patients are faring in the postoperative period. This suggests that a system to collate and visualize data for comparative feedback may allow anesthesiologists to fine-tune their decisions, optimizing pain and nausea outcomes.

As part of an organizational quality initiative, a postoperative follow-up (POFU) registry has been established at BC Children's Hospital, a tertiary pediatric hospital in Vancouver, British Columbia, Canada. Its purpose is to understand the recovery experience of day-case surgical patients and to facilitate associated quality improvement endeavors. The POFU registry is maintained by PACU clerks and nurses, who record day-surgery patient information and short-term outcomes from PACU and then follow-up with families via telephone to gather patient-reported outcomes at 24 hours post discharge. These data are recorded using the Research Electronic Data Capture (REDCap) web application (Vanderbilt University) [11,12] hosted locally. Each patient's contact information, demographics, clinician information, and procedure characteristics are entered using operating room scheduling system data. Validation checks are enabled for each field to ensure minimal artifacts; the data steward runs further reports and alerts to optimize data quality. Family-reported 24-hour outcomes (including postoperative pain, nausea, and vomiting) are collected by nurse telephone follow-up beginning in September 2020. PACU opioid and antiemetic administration, indicating early treatment of postoperative pain and nausea, have also been captured since April 2021.

This study's primary aim was to turn these outcome data into accessible and actionable information by creating dashboards, which allow the anesthesiologists to evaluate their patients' postoperative outcomes and reflect on their care. We also aimed to evaluate anesthesiologists' perception of the dashboard. By providing anesthesiologists with single-point access to these outcome data, allowing them to reflect on their care, drill down on details, and compare their performance to their peers and to the department aggregate in a time-efficient and user-friendly way, we aim to facilitate ongoing individual and departmental practices of improvement.

Methods

Study Design

We initially conducted a literature search of previous dashboard designs and partnered with a group of anesthesiologists in our department to identify their information needs and to co-design dashboards using an iterative development process. We then designed a dashboard architecture based on the POFU registry data; this incorporated key security features required to meet institutional policies and the confidentiality requirements of our anesthesiologists. We deployed the final design to the anesthesia department and conducted a preliminary usability evaluation.

Ethical Considerations

The University of British Columbia and Children's and Women's Health Centre of British Columbia Research Ethics Board determined this work to be established under a quality improvement or quality assurance (QIQA) framework (reviewed June 29, 2021), for which they do not require ethical review, in accordance with Article 2.5 of the Canadian Tri-Council Policy Statement. Data used in this study were obtained from the POFU registry, also established under a QIQA framework. This paper adheres to the Standards for Quality Improvement Reporting Excellence (version 2.0) guidelines [13].

Registry Data and Dashboard Architecture

Exploratory Data Analysis, Cleaning, and Processing

We performed an exploratory analysis in Python (Python Software Foundation) to confirm that the collected registry data were clean: variables were evaluated for any out-of-range values; anesthesiologist codes (departmental QIQA identifiers) and procedure codes were verified against lists of valid entries. Age, sex, and procedure group of PACU and missing 24-hour outcome data were compared to examine any significant differences in the underlying data. Cases with missing 24-hour outcomes or unanswered phone calls were excluded from the

XSL•FO
RenderX

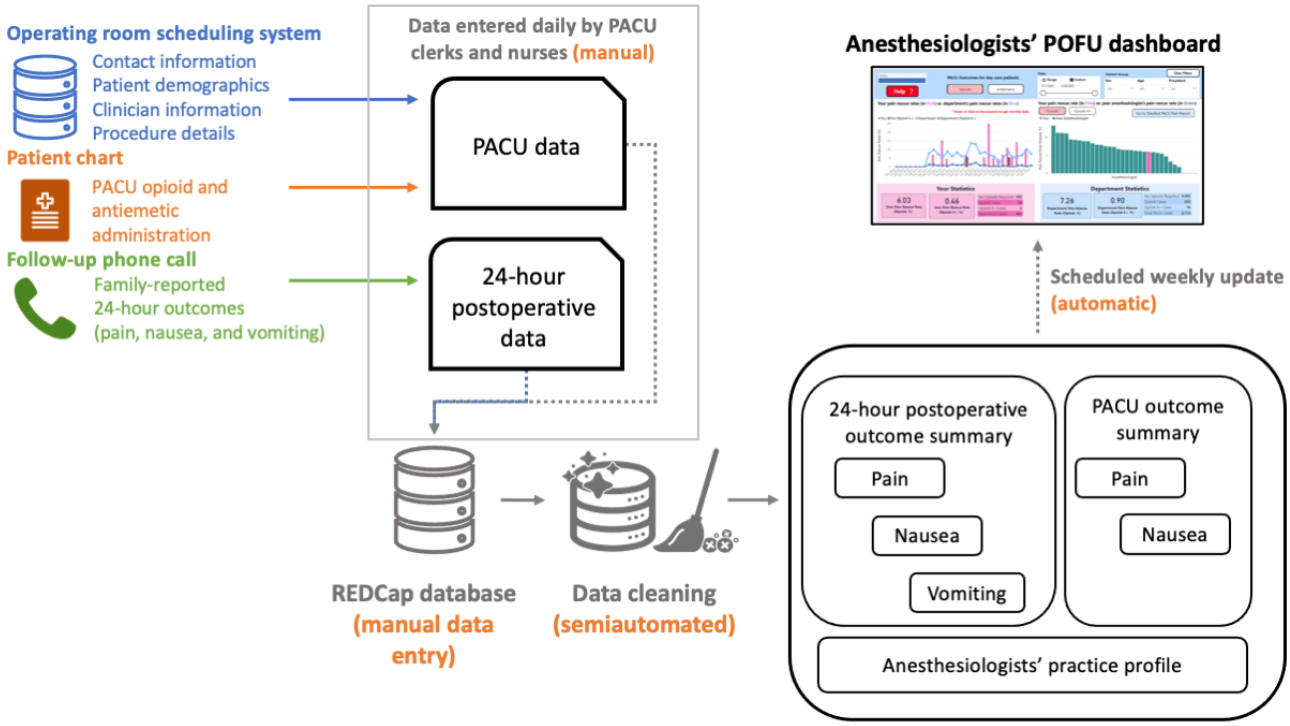
24-hour outcome analysis but not censured from the PACU outcome analysis.

Nausea and vomiting data were collected using a 4-point scale (none, mild, moderate, or severe). Pain scores in PACU are collected using developmentally appropriate observational or self-report tools, and 24-hour pain scores are obtained by proxy from a parent or caregiver using a 0-10 numeric rating scale [14]. We then assigned the same 4-point labels to the numeric pain ratings obtained: 0=none, 1-3=mild, 4-6=moderate, and 7-10=severe. For most dashboard purposes, we dichotomized variables into none or mild versus moderate or severe.

Dashboard Architecture and Data Sources

Our dashboard design aimed to provide a self-service platform for anesthesiologists to confidentially review their performance through postoperative (PACU and 24-hour) outcomes and anonymized peer comparison (Figure 1). The dashboards were developed using business analytics software Power BI Report Server (Microsoft), hosted by the BC Children's Hospital Research Institute (BCCHR). Power BI uses Active Directory Federation Services (Microsoft) to leverage hospital and research institute single sign-on and can, in principle, load data from REDCap, network drives, or other systems via application programming interfaces. At our institution, it is maintained by the BCCHR Data Management team, who manages access to team spaces in the server and determines access and usage policies.

Figure 1. The overview of the data flow from POFU data collection for PACU and 24-hour phone call into the POFU dashboard. PACU: postanesthetic care unit; POFU: postoperative follow-up.



Deidentified POFU registry data are downloaded weekly by the data steward as a comma separated values file and stored on a BCCHR secure data drive, that is, instead of employing an application programming interface token and directly linking these systems. BCCHR data management policies imposed the need for this workaround, as the POFU REDCap registry contains personal health information, parental contact information, and anesthesiologist and surgeon identifiers.

The dashboards require 3 files: the POFU weekly data export (excluding any identifying data), a Power BI username-QIQA code mapping file (see dashboard security and confidentiality), and an auxiliary file containing outcome definitions (severity level and numeric score mapping), procedure definitions (procedure code and group mapping), and age categories (National Institute of Child Health and Human Development pediatric age categories) [15].

The POFU registry and auxiliary files are set as automated data sources in Power BI, automatically refreshing the dashboard whenever files are updated. In contrast, the mapping file was added manually as a data source by the POFU data steward and cannot be downloaded by any other team member. Power BI pulls new data from these sources every Monday morning via a scheduled refresh.

The dashboard is available to anesthesiologists through a web browser and is typically accessed on a desktop or laptop computer; it can be viewed on a tablet or a smartphone, though it has not been optimized for use in this way.

Dashboard Security and Confidentiality

Comparative data are required to contextualize practice patterns for postoperative outcomes, but our team of co-design anesthesiologists insisted that the security and confidentiality of both patient and provider data were imperative for

XSL•FO
RenderX

departmental support: patient confidentiality is maintained by excluding identifiers from the data accessed by the dashboard; provider confidentiality is maintained through the use of departmental QIQA codes for anesthesiologists; see dashboard architecture and data sources. Data for the active anesthesiologist (ie, the user accessing the dashboard) and the department's aggregate performance are presented together, but access to the underlying data is restricted, preventing access to other anesthesiologists' data.

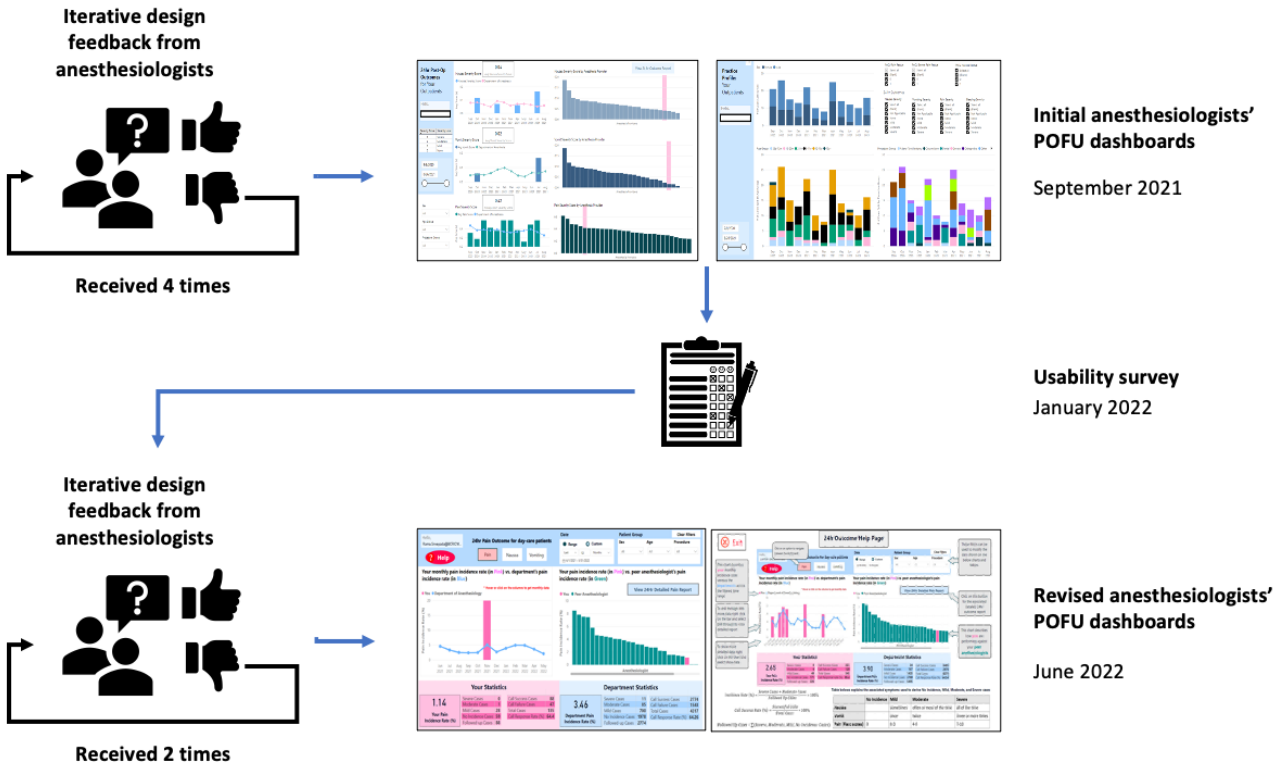
Power BI's row-level security feature filters the nonaggregated case data to include only data for cases the active anesthesiologist performed; hence, each user can only see their own cases. This is achieved via the Power BI username-QIQA code mapping. Access to this file is restricted to the POFU data steward, maintaining strict confidentiality of the anesthesiologists' identifiers.

Dashboard Design and Evaluation

Iterative Co-Design Sessions

We performed a brief literature search of papers published in 2015-2021 to understand existing anesthesia dashboards; keywords included *dashboard*, *run charts*, *personalized feedback*, *anesthesia feedback*, and *surgery feedback* [7,16-23]. Our preliminary visualization designs were based on the ideas drawn from this literature search. Guided by the recommendation to incorporate users' feedback into the dashboard development [16,19,23-26], we developed our system using a participatory design approach: we discussed design ideas with 5 anesthesiologists using iterative feedback to improve visualization. We adopted a convenience approach to selecting the design team, which consisted of the POFU clinical lead and other anesthesiologists who expressed an interest in contributing to the design process. Designs were demonstrated through screenshots and dynamic working prototypes during 6 feedback sessions; the visualizations were iteratively refined based on anesthesiologists' comments and observing their use of the prototype dashboards (Figure 2). Each iterative feedback meeting on Zoom (Zoom Video Communications Inc) lasted approximately 1 hour.

Figure 2. Iterative co-design process for the POFU dashboard. POFU: postoperative follow-up.



The initial dashboard prototypes showed a given anesthesiologist's average 24-hour postoperative pain, nausea, and vomiting scores and compared them with the department's average. Outcome scale distributions were also plotted. A second design added PACU outcomes (opioid and antiemetic administration). These were shown to the anesthesiologists, and the visualizations were iteratively refined to arrive at the final set of dashboards.

After the iterative development process, dashboards were deployed to the department in September 2021; the work was presented at departmental rounds, and anesthesiologists were emailed links to the tool with instructions and a contact email for further information. Subsequently, 2 further feedback sessions were conducted with our collaborating anesthesiologists to refine the final dashboard designs.

Usability Survey

To gather clinicians' feedback on the initial version of the dashboard, a usability survey was distributed to all department anesthesiologists in January 2022 after this an initial version of the dashboard had been available for 4 months. It consisted of

XSL•FO
RenderX

21 required and 18 optional follow-up questions on aspects such as frequency of use (including reason, if infrequent), ease of use (including clarity of information displayed and usefulness of the instructions), content (helpfulness and suggestions for other functionality), impact on practice, and overall opinion of the dashboard (Multimedia Appendix 1).

Based on the feedback received from the usability survey, the dashboards were redesigned and redeployed to the department in June 2022; the dashboards were again presented at department rounds, and anesthesiologist superusers were identified to provide training and peer support.

Dashboard Component Design

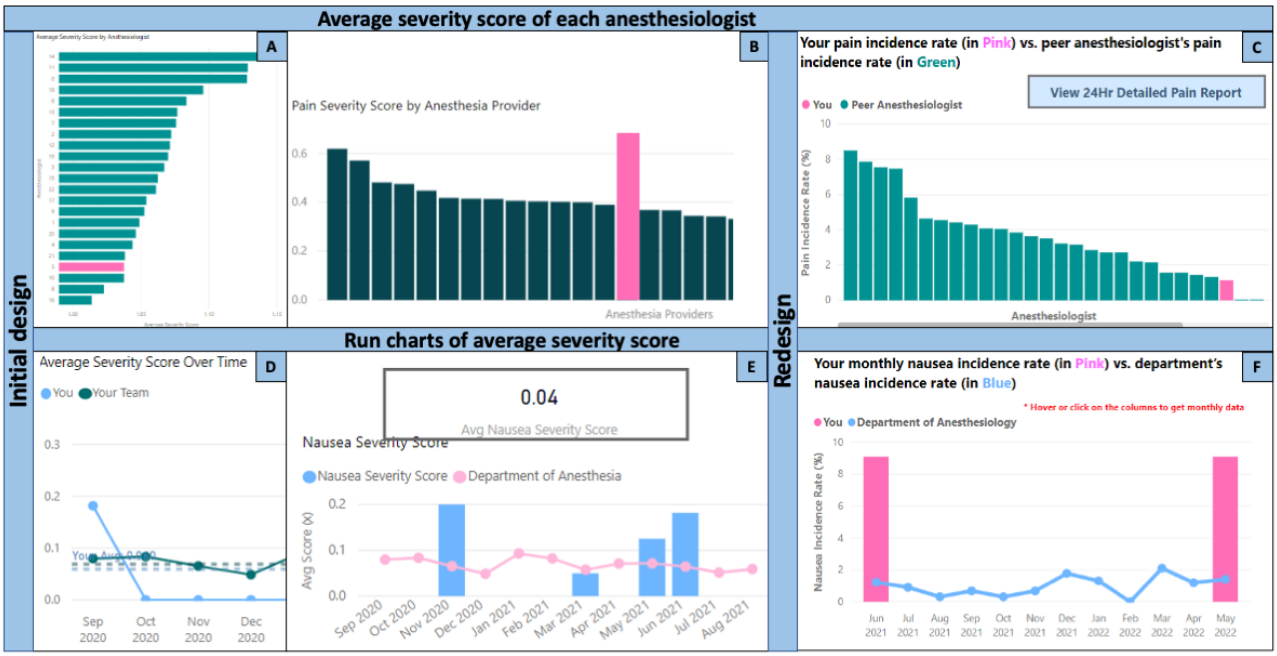
Overview

There are three dashboard components: (1) 24-hour outcome summary, (2) PACU outcome summary, and (3) anesthesiologists' practice profile.

24-Hour Outcome Summary

Guided by previous reports [19,24-26], 2 sections were developed initially: (1) average severity scores of patient outcomes [24,26], and (2) run charts of average severity score (Figures 3A and 3D). Inspired by Parks et al [24], we initially designed a bar chart that displayed the anesthesiologists' average severity scores in descending order, labeled with QIQA identifiers and the active anesthesiologist (user) highlighted (Figure 3A). Following feedback, we removed their codes to add a deidentification layer and added the severity score next to the bars (Figure 3B). The usability survey respondents indicated that the "pain severity score" calculation was unclear. Subsequently, we changed the metric from "pain severity score" (average of pain score categories) to "pain incidence rate" (occurrence of moderate or severe pain), which was more readily comprehensible (Figure 3C).

Figure 3. Progression of various sections of the dashboards with average severity scores in A, B, and pain incidence rates in C; monthly run charts of the anesthesiologist, along with team's aggregate outcomes in D, E, and F.



Based on co-design input, the active anesthesiologist's average severity score and the department's aggregate were initially represented in a monthly run chart with a median score superimposed (Figure 3D). The anesthesiologists subsequently suggested a run chart for the department-level score and a bar chart for the active anesthesiologist's score (Figure 3E). Finally, the title was modified to specify that the plot now reported incidence rates, and the colors were modified for consistency with other plots.

PACU Outcome Summary

The second phase of development added PACU outcome summary to the dashboards. PACU nurse interventions requiring opioid (both ≥1 dose and ≥4 doses) or antiemetic (≥1 dose) administration were plotted. The progression in design was similar to the 24-hour outcome dashboard. Monthly average PACU opioid or antiemetic administration rates of both the active anesthesiologist and the department were included in the final dashboards.

Anesthesiologists' Practice Profile

A practice profile page was designed based on anesthesiologists' interest in incorporating individual case mix information (age category, sex, and surgical service) into their interpretation of their outcome data.

Data Analysis

We summarized the 24-hour outcome data and PACU outcome data in the POFU database from the initial implementation of the data collection tool to illustrate the data available to anesthesiologists for visualization. We analyzed the anesthesiologists' usability survey responses and presented the results descriptively, along with the final dashboard designs. Finally, we conducted a preliminary analysis of the impact of the dashboards on postoperative patient outcomes by comparing

XSL•FO
RenderX

the baseline preimplementation period for PACU outcomes (April 2021 to June 2021) and 24-hour outcomes (September 2020 to June 2021) with the postimplementation period (September 2021 to February 2023) by plotting the month-by-month department incidence of each outcome and the overall average incidence for the period; to reduce bias, July 2021 to August 2021 have been plotted, but not considered in the comparison, as dashboard co-design was conducted during this period and was available to some members of the department. Changes between pre- and postimplementation periods were compared using the Wilcoxon rank-sum test.

Results

Data Set Characteristics

The 24-hour outcome dashboard contains data collected from September 2020 to February 2023. Of 12,082 total cases, 7877 (65.2%) postoperative phone calls were successfully completed to collect 24-hour outcomes from the family; 316 (4.0%) had moderate or severe pain, 73 (0.9%) had moderate or severe nausea, and 84 (1.1%) had moderate or severe vomiting. The distributions of age, sex, and PACU outcomes did not differ between patients with successful calls and those without. The PACU outcome dashboard contains 8716 cases collected from April 2021 to February 2023; 634 (7.3%) of these patients were administered at least 1 opioid dose, 78 (0.9%) required repeated opioid (≥4) doses, and 93 (1.1%) required an antiemetic.

Usability Survey

The January 2022 usability survey (Multimedia Appendix 1) was completed by 17 of 29 (59%) anesthesiologists, including 1 of 17 (6%) who had been practicing 6-11 years, 5 of 17 (29%) practicing 11-15 years, 6 of 17 (35%) practicing 16-20 years, and 5 of 17 (29%) practicing >20 years; respondents included 4 of 17 (24%) who performed cardiac anesthesia as part of their practice, 6 of 17 (35%) who performed anesthesia for neurosurgery, and 4 of 17 (24%) who performed anesthesia for spine surgery. Among all respondents, 15 of 17 had used the dashboard during the 4 months since the initial version had been deployed, though only 2 of 17 (12%) had used it regularly. Of those who had used the dashboard, 9 of 15 (60%) reported that it was easy to navigate, and 9 of 15 (60%) thought the information was clearly presented. On the other hand, 3 of 15 (20%) users had found the navigation difficult, 2 of 15 (13%) thought the information needed to be clearer, and only 7 of 15 (47%) had found the help text and user instructions helpful.

Overall, the information provided was considered helpful by 12 of 15 (80%) of those that had used the dashboard, some of whom indicated that it had impacted their practice: for 8 of 15 (53%), this impact had been minimal, but 2 of 15 (13%) considered its impact significant. Comments indicated that the perceived benefits were primarily related to an increased awareness of the need for higher intraoperative analgesic and antiemetic dosing in some cases and an opportunity to engage with trainees on this issue. Most respondents (12/17, 71%) confirmed they were comfortable with how this information about their practice was being collected and presented to them, although 4 of 17 (24%) were concerned there may be negative consequences to having these data available. Concerns were that the quality indicators presented did not consider the multiple other influences on patient outcomes and, depending on how the indicators are subsequently used, could apportion blame inappropriately, with possible professional or legal consequences; this further highlights the need to guarantee provider confidentiality.

Concerning future use, 14 of 17 (82%) respondents recommended a regular reminder email, and 11 of 17 (64%) indicated they would refer to the dashboard at least monthly, with 10 of 17 (59%) believing that it had the potential to have a significant benefit for their practice. Additional information requirements identified for future work included: duration of PACU stay, 13 of 17 (76%); antibiotics timing, 8 of 17 (47%); perioperative hypothermia, 8 of 17 (47%); difficult intubation recorded, 6 of 17 (35%); and incidence of hypotension, 5 of 17 (29%).

Final Dashboards Deployed

Overview

The final dashboard components, deployed in June 2022, are (1) 24-hour outcome summary (Figure 4), (2) PACU outcome summary (Figure 5), and (3) anesthesiologists' practice profile (Figure 6).

XSL•FO
RenderX

Figure 4. The final 24-hour postoperative outcome summary dashboard with notes.

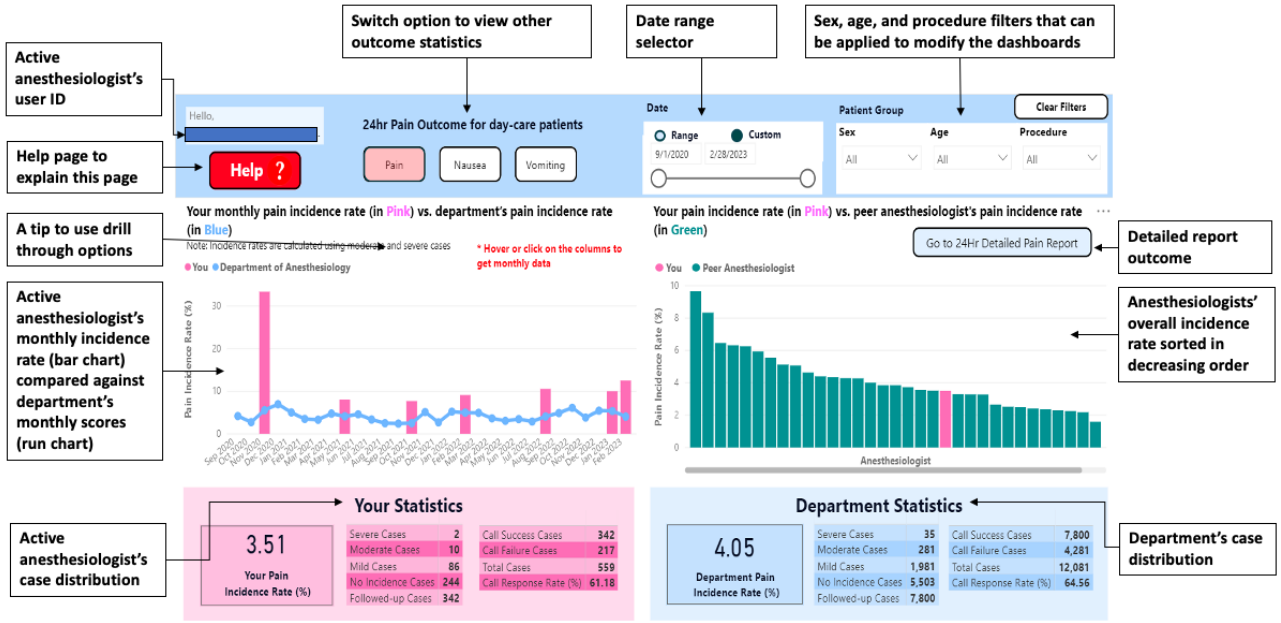
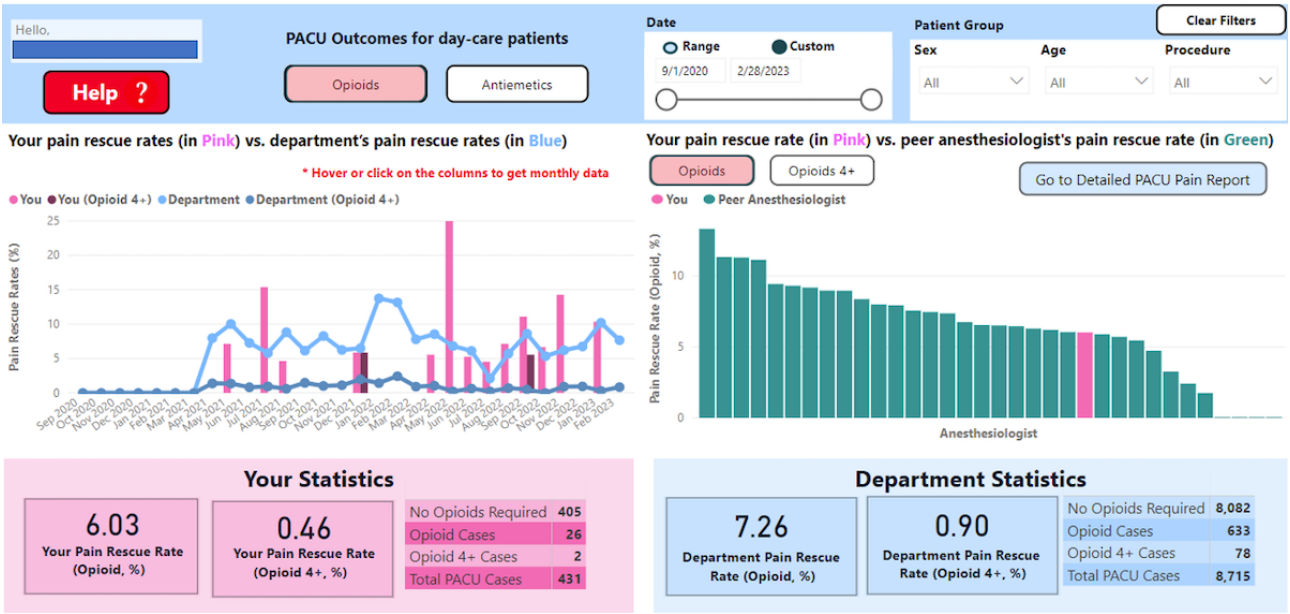
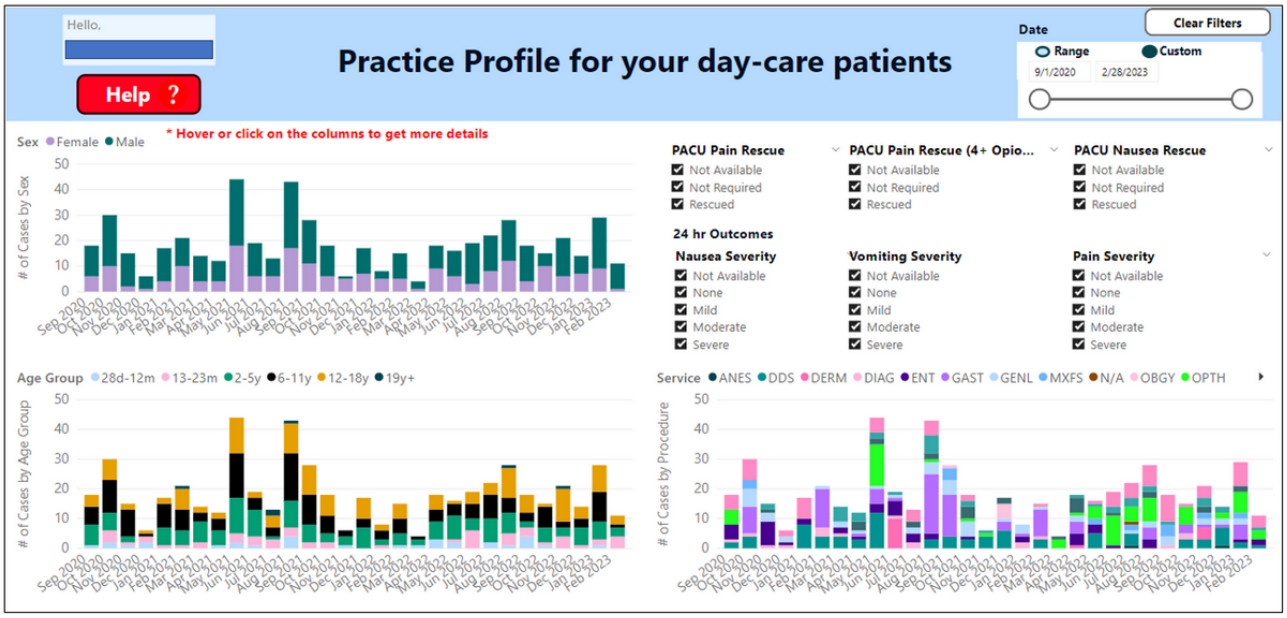


Figure 5. The final PACU outcomes dashboard design. PACU: postanesthetic care unit.



XSL•FO
RenderX

Figure 6. The final anesthesiologist practice profile dashboard design.



24-Hour Outcome Summary

The 24-hour outcome summary displays each outcome (pain, nausea, and vomiting) and compares the active anesthesiologist's monthly incidence rate with the department's rate (Figure 4, left). Their peers' incidence rates are given in descending order, with the active anesthesiologist highlighted (Figure 4, right). Guided by the usability survey, we added the active anesthesiologist's and overall department statistics (Figure 4, bottom).

The *View 24hr Outcome Report* button leads to a detailed report, where the anesthesiologist can view additional outcome information via switch tabs at the top of the page (Figure S1 in Multimedia Appendix 2): this report displays the anesthesiologist's case distribution across the 4 severity levels and the outcome incidence rate by sex, age category, and procedure group. All data are shown compared to the department's rates.

Both the outcome summary and the detailed report, support filtering based on patient sex, age category, and surgical service using drop-down menus. The "show data" option lets users view deidentified data contained within each plot. For a patient-level or filtered view of an anesthesiologist's caseload, the user can apply a "drill-through" feature on all the personal charts to dive deeper into the data using advanced filters.

Based on the usability survey, we reconfigured the user instruction manual as a help page (Figure S2 in Multimedia Appendix 2), with a screenshot of the outcome page and information about each section, including formulas to calculate incidence rates.

PACU Outcome Summary

The PACU outcome summary uses the same visualization techniques as the 24-hour outcome summary: (1) plot for overall change over time and (2) for peer comparison, the opioid rescue rate is further divided into pain that required at least 1 dose and pain that required ≥4 doses of rescue medication (opioids 4+).

The PACU outcome summary shows opioid and antiemetic administration rates on separate pages: *PACU opioids* (Figure 5) and *PACU antiemetics* (Figure S3 in Multimedia Appendix 2). A toggle switch between opioids and opioids 4+ allows the anesthesiologists to view their performance against their peers (Figure 5, right). This PACU summary has the same help and data management capabilities as the 24-hour outcomes dashboard.

Final Anesthesiologists' Practice Profile

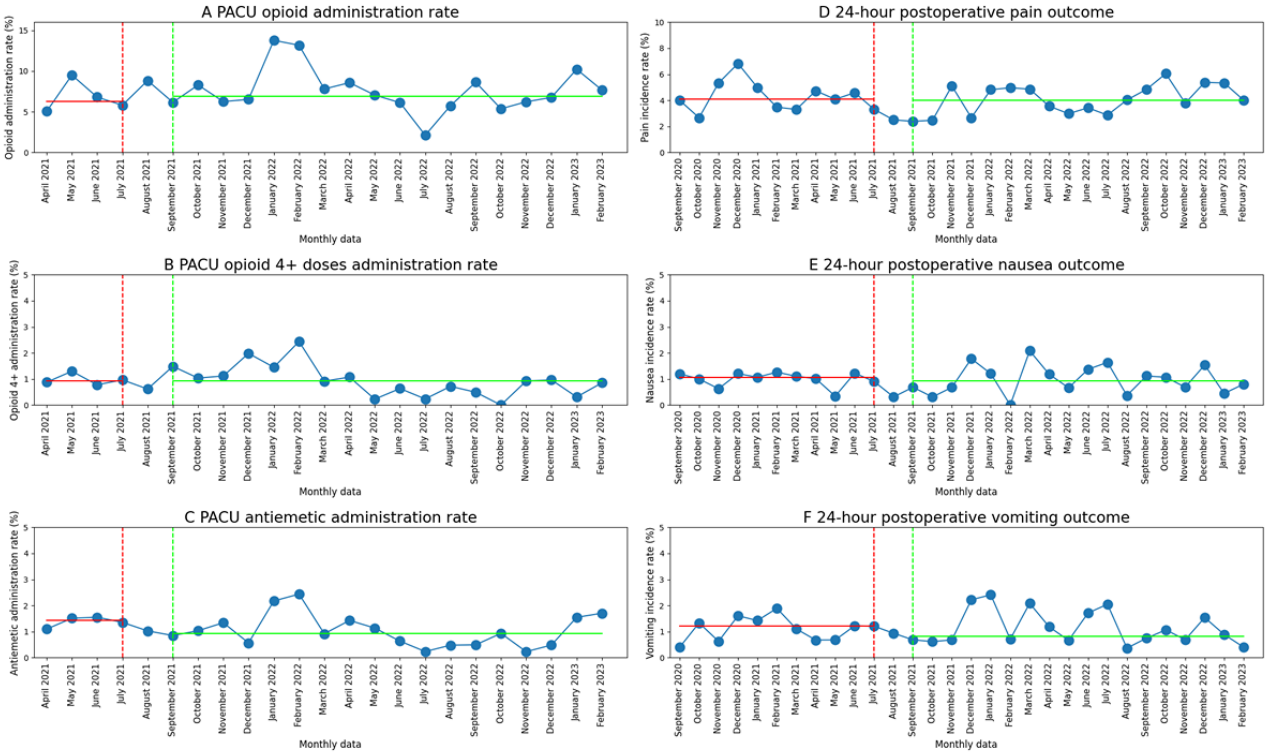
A practice profile allows the anesthesiologist to view their case mix in stacked bar plots grouped by sex, age category, and surgical service in monthly intervals (Figure 6). The practice profile supports filtering the anesthesiologist's case mix based on PACU and 24-hour postoperative outcomes. The user can also view the average outcomes of each caseload group as a tool tip or "drill-through" to see a detailed list of cases.

Preliminary Analysis of Dashboard Impact

The department's aggregate incidence rate for all outcomes is reasonably low but with significant month to month variability (Figure 7). Incidence rates were not different between the preimplementation and postimplementation periods: median differences were 0.7 (95% CI –2.5 to 3.4; P=.52) for PACU opioid administration, –0.0 (95% CI –0.7 to 0.5; P=.76) for PACU opioids 4+ administration, –0.5 (95% CI –1.0 to 0.2; P=.16) for PACU antiemetic administration, –0.1 (95% CI –1.2 to 0.8; P=.86) for 24-hour pain, –0.0 (95% CI –0.4 to 0.3; P=.88) for 24-hour nausea, and 0.0 (95% CI –0.5 to 0.5; P>.99) for 24-hour vomiting.

XSL•FO
RenderX

Figure 7. Variability in the department's monthly aggregate PACU outcome rates, including (A) opioid rescue, (B) opioid rescue ≥4 doses, and (C) antiemetic administration, and 24-hour postoperative outcome rates, including (D) pain, (E) nausea, and (F) vomiting. The department's monthly aggregate outcome rates are plotted as blue dots; the red dotted line indicates the end of the preimplementation period; the green dotted line indicates the start of the postimplementation period; the solid red and green horizontal lines indicate the medians of the pre- and postimplementation periods, respectively. PACU: postanesthetic care unit.



Discussion

Principal Results

We designed and implemented dynamic dashboards for the anesthesiologists in our department to review their outpatients' postoperative short-term outcomes and compare these outcomes to their peers to facilitate personal performance evaluation and potential practice improvements. This project built on previous work establishing the POFU registry to collect and electronically aggregate these outcome data. The dashboard is updated weekly and presents (1) 24-hour outcomes, (2) PACU outcomes, and (3) a practice profile containing individual anesthesiologist's case mix, grouped by age groups, sex, and surgical service. Co-design and usability evaluation indicated requirements for uncluttered summary data visualization and the ability to explore specific outcome details if needed. Anesthesiologists in our department generally found the dashboard information helpful and believed it would significantly benefit their practice. Our study confirmed that ensuring security and patient and provider confidentiality in these dashboards was crucial for successful uptake.

Comparison With Prior Work

There are significant limitations in using postoperative outcomes such as pain, nausea, and vomiting as anesthesiologist performance indicators. There are many confounding factors, including procedural details and the surgeon being responsible for prescribing analgesics. These confounders will likely increase with increasing time after the procedure. However, even for immediate postanesthesia outcomes, confounders, including the PACU nursing team assessing and managing pain, have been shown to invalidate interanesthesiologist performance comparisons [27]. It is also essential to consider the impact of the case mix: Schulz et al [28] suggest that case-mix adjustment of measures such as length of PACU stay may provide more meaningful indicators than unadjusted metrics. However, the dashboard was not founded on the idea that an anesthesiologist's practice is the sole determinant of their patients' outcomes; it is a tool for anesthesiologists to reflect on their practice in context. Providing the data is only the first step toward better understanding their practices, but it is necessary. Anesthesiologists know that multiple variables outside their control contribute to these data, especially for 24-hour outcomes. Understanding their patients' variability should allow them to interpret this information appropriately.

Clinical dashboards are being widely explored in an attempt to better understand and optimize patient outcomes in a range of anesthesia settings [29], including cardiac [30] and pediatric anesthesia [18,20,31]. Data analytics focus on process, as well as outcome, metrics [32], and there is often a need to extract the required data from a range of in-hospital systems [31] or to supplement these data with patient- or family-reported outcomes [20,33]. These initiatives may or may not have a positive impact on performance. For example, regular team and individualized feedback reduced temperature monitoring delays during spine surgery among 1 group of anesthesiologists [7]. In contrast, another group found that audit and feedback did not improve the intraoperative temperature management [34]. However, such

XSL•FO
RenderX

initiatives may often be the best route to improving practice, patient experience, and outcomes, particularly in acute pain management, given the challenges faced in implementing and adopting integrated electronic medical records in this area [35].

While dashboards and follow-up phone calls may function as intended, be economically justifiable, and be well received by anesthesiologists [20,31,36], it will be more challenging to demonstrate a beneficial impact on patient-relevant outcomes rather than process outcomes. Brenn et al [20] reported an 88% response rate for postoperative phone calls to 42,688 pediatric outpatients, but were unable to link satisfaction with complication rates and suggested that reducing wait times and streamlining operations are more important to families. In contrast, Kim et al [33] found that implementing a follow-up call (even from a nonmedical professional) 48 hours after day surgery reduced family anxiety, though it did not improve family satisfaction. We have yet to examine the effect of our dashboard on improving our patients' outcomes.

Through an iterative design process, we learned the importance of collaborating with the end users of the dashboards to maximize usability. This collaboration in the co-design process familiarized us with local clinical terminology (eg, “day-care patients” in preference to “ambulatory patients” or “outpatients”), which helped us design dashboards that are more locally relevant. These collaborative design processes have been adopted by other researcher-developers [18,19,23,31,33], although not all such dashboards have been co-designed with end users or undergone usability evaluation [17,30]. We used a usability survey to guide us toward more acceptable dynamic dashboards with various analysis options.

Limitations

Some limitations of our dashboard designs and development approach should be noted. First, there is no risk adjustment of performance scores based on case mix: an anesthesiologist with a significant proportion of patients in whom postoperative pain or nausea have a higher baseline incidence cannot be validly compared to an anesthesiologist with a different case mix. The dashboard users recognize this limitation, and we implemented the practice profile section partly in an effort to address this issue. We may explore risk adjustment strategies in future but must be cautious not to overturn our users’ requirement for concise summary performance reports.

Second, the POFU registry includes age, sex, surgical service, and procedure date but does not contain other presurgical (risk) factors such as weight, comorbidities, existing medication, allergies, number or type of previous procedures, or child and parent anxiety levels. In the future, we plan to add additional patient data, surgical and anesthetic techniques, and PACU length of stay by integrating with the hospital’s Anesthesia Information Management System (SurgiNet Anesthesia, Cerner), which may allow us to provide our anesthesiologists with further insights.

Third, our current data are limited to outcomes occurring within 24 hours of discharge: PACU data recorded according to institutional practice and the 24-hour outcome data, for which a nurse makes only a single phone call to the patient’s family. The success rate to date has only been 7800 out of 12,081 (65%) cases; a higher rate would improve validity. To alleviate the missing 24-hour outcome data, we are exploring the development of self-reporting tools for pain, nausea, and vomiting, such as Panda [37], a mobile postoperative pain management app. We also aim to determine if families are willing to provide outcomes beyond 24 hours.

Finally, our usability questionnaire was not a standardized instrument or evaluated for reliability or validity, which limits its generalizability. It aimed to evaluate specific dashboard features as a final step before department-wide deployment.

Further Evaluation

The dashboard version has been deployed and is being used by department anesthesiologists, with data updated weekly. We will perform an ongoing usability evaluation to examine the dashboards' usefulness, determine if anesthesiologists have any issues with the visualizations, and explore suggestions for additional information or functionality, including further optimization for use on a tablet or smartphone if required. We plan to evaluate usage patterns: how many anesthesiologists use the dashboards, how frequently, for what purpose, and over what period. Finally, we aim to extend our analysis of changes in the department’s aggregate and individual outcomes post deployment. Our preliminary analysis did not demonstrate any significant changes in outcomes, which may be in part because our PACU and 24-hour follow-up outcomes were already reasonably well-optimized compared to other institutions’ [38,39]. Our future work will focus on tracking and reducing the variability in outcome rates between different practitioners. Ultimately, evaluating the impact of this initiative on our patients’ outcomes is a significant undertaking and, hence, a long-term goal. It will involve integrating with our Anesthesia Information Management System (SurgiNet Anesthesia, Cerner), identifying other key outcomes that matter to our patients and their families, and applying the appropriate systems and resources to collect, process, and analyze this information.

Conclusions

The dynamic, personalized dashboards we designed and deployed have allowed the anesthesiologists in our department to review their outpatients’ short-term postoperative outcomes and reflect on their practice in the context of peer and departmental performance. Key lessons from this implementation include the value of adopting a participatory approach to development, with co-design workshops and usability evaluation; the importance of establishing a robust approach to the security and confidentiality of patient and provider data in gaining user trust; and the preference for presenting uncluttered summary data combined with the opportunity to drill down into specific cases if required. This dashboard solution has allowed our department’s anesthesiologists to visualize previously unavailable data collected as part of a broader quality initiative. It should provide the opportunity to evaluate iterative practice changes, although further work will be required to monitor its effect on individual and department performance and patient outcomes.

XSL•FO
RenderX

Acknowledgments

Data reported in this study were retrieved from the BC Children's Hospital postoperative follow-up (POFU) database. The authors would like to thank the POFU registry collaborators: the Pediatric Anesthesia Research Team, BC Children's Hospital (BCCH) Procedures and Surgical Services, and BCCH Surgery and Anesthesia, for collecting the data and granting us access to them for this work. We thank the BC Children's Hospital Foundation for its support. This work was supported, in part, by a UBC Canada's Digital Technology Supercluster expansion research grant. The POFU registry was supported by a Medical Allied Staff Engagement Society Engagement Fund grant (awarded to NRF) and a BCCH Surgical Program Improvement Fund grant (awarded to NRF and Elizabeth Lamb). MG holds a Michael Smith Health Research BC scholar award; RSS held a Mitacs postdoctoral fellowship while this work was conducted. This work was previously presented, in part, at the Society for Technology in Anesthesia 2022 Virtual Annual Meeting, January 2022 and the Canadian Pediatric Anesthesia Society Annual Meeting in Vancouver in September 2022.

Conflicts of Interest

None declared.

Multimedia Appendix 1
POFU Anesthesia Dashboard - usability and utility survey.
[PDF File (Adobe PDF File), 89 KB - periop_v6i1e47398_app1.pdf]

Multimedia Appendix 2
View 24hr Outcome Report, Help page, and PACU Antiemetics.
[PDF File (Adobe PDF File), 530 KB - periop_v6i1e47398_app2.pdf]

References

1. Benn J, Arnold G, Wei I, Riley C, Aleva F. Using quality indicators in anaesthesia: feeding back data to improve care. Br J Anaesth 2012;109(1):80-91 [FREE Full text] [doi: 10.1093/bja/aes173] [Medline: 22661749]
2. Benn J, Arnold G, D'Lima D, Wei I, Moore J, Aleva F, et al. Evaluation of a continuous monitoring and feedback initiative to improve quality of anaesthetic care: a mixed-methods quasi-experimental study. In: Heal Serv Deliv Res. Southampton (UK): NIHR Journals Library; 2015.
3. Nair BG, Newman SF, Peterson GN, Wu WY, Schwid HA. Feedback mechanisms including real-time electronic alerts to achieve near 100% timely prophylactic antibiotic administration in surgical cases. Anesth Analg 2010;111(5):1293-1300. [doi: 10.1213/ANE.0b013e3181f46d89] [Medline: 20841414]
4. Sutherland T, Beloff J, Lightowler M, Liu X, Nascimben L, Kaye AD, et al. Description of a multidisciplinary initiative to improve SCIP measures related to pre-operative antibiotic prophylaxis compliance: a single-center success story. Patient Saf Surg 2014;8(1):37 [FREE Full text] [doi: 10.1186/s13037-014-0037-2] [Medline: 25431623]
5. Malapero RJ, Gabriel RA, Gimlich R, Ehrenfeld JM, Philip BK, Bates DW, et al. An anesthesia medication cost scorecard—concepts for individualized feedback. J Med Syst 2015;39(5):48. [doi: 10.1007/s10916-015-0226-2] [Medline: 25732076]
6. Gabriel RA, Gimlich R, Ehrenfeld JM, Urman RD. Operating room metrics score card-creating a prototype for individualized feedback. J Med Syst 2014;38(11):144. [doi: 10.1007/s10916-014-0144-8] [Medline: 25315824]
7. Görges M, West NC, Whyte SD. Using physiological monitoring data for performance feedback: an initiative using thermoregulation metrics. Can J Anaesth 2017;64(3):245-251. [doi: 10.1007/s12630-016-0762-3] [Medline: 27785761]
8. D'Lima DM, Moore J, Bottle A, Brett SJ, Arnold GM, Benn J. Developing effective feedback on quality of anaesthetic care: what are its most valuable characteristics from a clinical perspective? J Health Serv Res Policy 2015;20(1 Suppl):26-34 [FREE Full text] [doi: 10.1177/1355819614557299] [Medline: 25472987]
9. Epstein RH, Dexter F, Patel N. Influencing anesthesia provider behavior using anesthesia information management system data for near real-time alerts and post hoc reports. Anesth Analg 2015;121(3):678-692. [doi: 10.1213/ANE.0000000000000677] [Medline: 26262500]
10. Coté CJ, Lerman J, Anderson BJ. The practice of pediatric anesthesia. In: A Practice of Anesthesia for Infants and Children. Philadelphia: Elsevier; 2019:1-7.
11. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]

XSL•FO
RenderX

12. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
13. Ogrinc G, Davies L, Goodman D, Batalden P, Davidoff F, Stevens D. SQUIRE 2.0 (Standards for QUality Improvement Reporting Excellence): revised publication guidelines from a detailed consensus process. BMJ Qual Saf 2016;25(12):986-992 [FREE Full text] [doi: 10.1136/bmjqs-2015-004411] [Medline: 26369893]
14. Birnie KA, Hundert AS, Lalloo C, Nguyen C, Stinson JN. Recommendations for selection of self-report pain intensity measures in children and adolescents: a systematic review and quality assessment of measurement properties. Pain 2019;160(1):5-18. [doi: 10.1097/j.pain.0000000000001377] [Medline: 30180088]
15. Williams K, Thomson D, Seto I, Contopoulos-Ioannidis DG, Ioannidis JPA, Curtis S, StaR Child Health Group. Standard 6: age groups for pediatric trials. Pediatrics 2012;129(Suppl 3):S153-S160. [doi: 10.1542/peds.2012-0055I] [Medline: 22661762]
16. Elshehaly M, Randell R, Brehmer M, McVey L, Alvarado N, Gale CP, et al. QualDash: adaptable generation of visualisation dashboards for healthcare quality improvement. IEEE Trans Vis Comput Graph 2021;27(2):689-699 [FREE Full text] [doi: 10.1109/TVCG.2020.3030424] [Medline: 33048727]
17. Reilly CA, Doughty HP, Werth PM, Rockwell CW, Sparks MB, Jevsevar DS. Creating a value dashboard for orthopaedic surgical procedures. J Bone Joint Surg Am 2020;102(21):1849-1856. [doi: 10.2106/JBJS.20.00270] [Medline: 32694400]
18. Nelson O, Sturgis B, Gilbert K, Henry E, Clegg K, Tan JM, et al. A visual analytics dashboard to summarize serial anesthesia records in pediatric radiation treatment. Appl Clin Inform 2019;10(4):563-569 [FREE Full text] [doi: 10.1055/s-0039-1693712] [Medline: 31390667]
19. Patel MS, Kurtzman GW, Kannan S, Small DS, Morris A, Honeywell S, et al. Effect of an automated patient dashboard using active choice and peer comparison performance feedback to physicians on statin prescribing: the PRESCRIBE cluster randomized clinical trial. JAMA Netw Open 2018;1(3):e180818 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.0818] [Medline: 30646039]
20. Brenn BR, Choudhry DK, Sacks K. Outpatient outcomes and satisfaction in pediatric population: data from the postoperative phone call. Paediatr Anaesth 2016;26(2):158-163. [doi: 10.1111/pan.12817] [Medline: 26612631]
21. O'Reilly-Shah VN, Easton GS, Jabaley CS, Lynde GC. Variable effectiveness of stepwise implementation of nudge-type interventions to improve provider compliance with intraoperative low tidal volume ventilation. BMJ Qual Saf 2018;27(12):1008-1018. [doi: 10.1136/bmjqs-2017-007684] [Medline: 29776982]
22. Dowding D, Randell R, Gardner P, Fitzpatrick G, Dykes P, Favela J, et al. Dashboards for improving patient care: review of the literature. Int J Med Inform 2015;84(2):87-100 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.10.001] [Medline: 25453274]
23. Wu DTY, Vennemeyer S, Brown K, Revalee J, Murdock P, Salomone S, et al. Usability testing of an interactive dashboard for surgical quality improvement in a large congenital heart center. Appl Clin Inform 2019;10(5):859-869 [FREE Full text] [doi: 10.1055/s-0039-1698466] [Medline: 31724143]
24. Parks DA, Short RT, McArdle PJ, Liwo A, Hagood JM, Crump SJ, et al. Improving adherence to intraoperative lung-protective ventilation strategies using near real-time feedback and individualized electronic reporting. Anesth Analg 2021;132(5):1438-1449. [doi: 10.1213/ANE.0000000000005481] [Medline: 33724961]
25. McCormick PJ, Yeoh C, Vicario-Feliciano RM, Ervin K, Tan KS, Yang G, et al. Improved compliance with anesthesia quality measures after implementation of automated monthly feedback. J Oncol Pract 2019;15(6):e583-e592 [FREE Full text] [doi: 10.1200/JOP.18.00521] [Medline: 31107625]
26. Bowdle TA, Jelacic S, Nair B, Zucker F, Bussey LS, Togashi K, et al. Electronic audit and feedback with positive rewards improve anesthesia provider compliance with a barcode-based drug safety system. Anesth Analg 2019;129(2):418-425. [doi: 10.1213/ANE.0000000000003861] [Medline: 30320650]
27. Wanderer JP, Shi Y, Schildcrout JS, Ehrenfeld JM, Epstein RH. Supervising anesthesiologists cannot be effectively compared according to their patients' postanesthesia care unit admission pain scores. Anesth Analg 2015;120(4):923-932. [doi: 10.1213/ANE.0000000000000480] [Medline: 25325754]
28. Schulz EB, Phillips F, Waterbright S. Case-mix adjusted postanaesthesia care unit length of stay and business intelligence dashboards for feedback to anaesthetists. Br J Anaesth 2020;125(6):1079-1087 [FREE Full text] [doi: 10.1016/j.bja.2020.06.068] [Medline: 32863015]
29. Barbeito A, Segall N. Development and usability testing of an audit and feedback tool for anesthesiologists. JAMIA Open 2019;2(1):29-34 [FREE Full text] [doi: 10.1093/jamiaopen/ooy054] [Medline: 31984342]
30. Hensley NB, Grant MC, Cho BC, Suffredini G, Abernathy JA. How do we use dashboards to enhance quality in cardiac anesthesia? J Cardiothorac Vasc Anesth 2021;35(10):2969-2976. [doi: 10.1053/j.jvca.2021.04.034] [Medline: 34059439]
31. Cravero JP, Sriswasdi P, Lekowski R, Carpino E, Blum R, Askins N, et al. Creation of an integrated outcome database for pediatric anesthesia. Paediatr Anaesth 2016;26(4):345-355. [doi: 10.1111/pan.12857] [Medline: 26956515]
32. Valentine EA, Falk SA. Quality improvement in anesthesiology—leveraging data and analytics to optimize outcomes. Anesthesiol Clin 2018;36(1):31-44. [doi: 10.1016/j.anclin.2017.10.006] [Medline: 29425597]

XSL•FO
RenderX

33. Kim JK, Lee MJ, Chua ME, Ming JM, Lorenzo AJ, Farhat WA, et al. Do post-operative phone calls enhance family satisfaction and outcomes after outpatient pediatric urological surgeries? A prospective study. Pediatr Surg Int 2021;37(1):161-167. [doi: 10.1007/s00383-020-04770-5] [Medline: 33136281]
34. Boet S, Bryson GL, Taljaard M, Pigford AA, McIsaac DI, Brehaut J, Canadian Perioperative Anesthesia Clinical Trials Group. Effect of audit and feedback on physicians' intraoperative temperature management and patient outcomes: a three-arm cluster randomized-controlled trial comparing benchmarked and ranked feedback. Can J Anaesth 2018;65(11):1196-1209. [doi: 10.1007/s12630-018-1205-0] [Medline: 30159716]
35. Goldstein DH, Phelan R, Wilson R, Ross-White A, VanDenKerkhof EG, Penning JP, et al. Brief review: Adoption of electronic medical records to enhance acute pain management. Can J Anaesth 2014;61(2):164-179. [doi: 10.1007/s12630-013-0069-6] [Medline: 24233770]
36. Laurent G, Moussa MD, Cirenei C, Tavernier B, Marcilly R, Lamer A. Development, implementation and preliminary evaluation of clinical dashboards in a department of anesthesia. J Clin Monit Comput 2021;35(3):617-626 [FREE Full text] [doi: 10.1007/s10877-020-00522-x] [Medline: 32418147]
37. Dunsmuir D, Wu H, Sun T, West NC, Lauder GR, Görges M, et al. A postoperative pain management mobile app (panda) for children at home after discharge: usability and feasibility. JMIR Perioper Med 2019;2(2):e12305 [FREE Full text] [doi: 10.2196/12305] [Medline: 33393928]
38. Ames WA, Machovec K. An update on the management of PONV in a pediatric patient. Best Pract Res Clin Anaesthesiol 2020;34(4):749-758. [doi: 10.1016/j.bpa.2020.05.007] [Medline: 33288124]
39. Andrew BY, Habib AS, Taicher BM. The association of guideline-directed prophylaxis with postoperative nausea and vomiting in pediatric patients: a single-center, retrospective cohort study. Anesth Analg 2023:preprint. [doi: 10.1213/ANE.0000000000006543] [Medline: 37235529]

Abbreviations

BCCHR: British Columbia Children's Hospital Research Institute
PACU: postanesthetic care unit
POFU: postoperative follow-up registry
QIQA: quality improvement or quality assurance
REDCap: Research Electronic Data Capture

Edited by T de Azevedo Cardoso; submitted 18.03.23; peer-reviewed by C Matava, J Galvez, N Alvarado; comments to author 15.06.23; revised version received 08.08.23; accepted 16.08.23; published 19.09.23.

Please cite as:
Sreepada RS, Chang AC, West NC, Sujan J, Lai B, Poznikoff AK, Munk R, Froese NR, Chen JC, Görges M
Dashboard of Short-Term Postoperative Patient Outcomes for Anesthesiologists: Development and Preliminary Evaluation
JMIR Perioper Med 2023;6:e47398
URL: https://periop.jmir.org/2023/1/e47398
doi:10.2196/47398
PMID:37725426

©Rama Syamala Sreepada, Ai Ching Chang, Nicholas C West, Jonath Sujan, Brendan Lai, Andrew K Poznikoff, Rebecca Munk, Norbert R Froese, James C Chen, Matthias Görges. Originally published in JMIR Perioperative Medicine (http://periop.jmir.org), 19.09.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Perioperative Medicine, is properly cited. The complete bibliographic information, a link to the original publication on http://periop.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

An Ensemble Learning Approach to Improving Prediction of Case Duration for Spine Surgery: Algorithm Development and Validation

Rodney Allanigue Gabriel[1,2], MD, MAS; Bhavya Harjai[2], BS; Sierra Simpson[2], PhD; Austin Liu Du[3], BS; Jeffrey Logan Tully[2], MD; Olivier George[4], PhD; Ruth Waterman[5], MD

[1]Division of Biomedical Informatics, Department of Medicine, University of California, San Diego, San Diego, CA, United States
[2]Division of Perioperative Informatics, Department of Anesthesiology, University of California, San Diego, San Diego, CA, United States
[3]School of Medicine, University of California, San Diego, San Diego, CA, United States
[4]Department of Psychiatry, University of California, San Diego, San Diego, CA, United States
[5]Department of Anesthesiology, University of California, San Diego, San Diego, CA, United States

Corresponding Author:
Rodney Allanigue Gabriel, MD, MAS
Division of Biomedical Informatics
Department of Medicine
University of California, San Diego
9300 Campus Point Dr
San Diego, CA, 92037
United States
Phone: 1 858 657 7000
Email: ragabriel@health.ucsd.edu

Abstract

Background: Estimating surgical case duration accurately is an important operating room efficiency metric. Current predictive techniques in spine surgery include less sophisticated approaches such as classical multivariable statistical models. Machine learning approaches have been used to predict outcomes such as length of stay and time returning to normal work, but have not been focused on case duration.

Objective: The primary objective of this 4-year, single-academic-center, retrospective study was to use an ensemble learning approach that may improve the accuracy of scheduled case duration for spine surgery. The primary outcome measure was case duration.

Methods: We compared machine learning models using surgical and patient features to our institutional method, which used historic averages and surgeon adjustments as needed. We implemented multivariable linear regression, random forest, bagging, and XGBoost (Extreme Gradient Boosting) and calculated the average R^2, root-mean-square error (RMSE), explained variance, and mean absolute error (MAE) using k-fold cross-validation. We then used the SHAP (Shapley Additive Explanations) explainer model to determine feature importance.

Results: A total of 3189 patients who underwent spine surgery were included. The institution's current method of predicting case times has a very poor coefficient of determination with actual times (R^2=0.213). On k-fold cross-validation, the linear regression model had an explained variance score of 0.345, an R^2 of 0.34, an RMSE of 162.84 minutes, and an MAE of 127.22 minutes. Among all models, the XGBoost regressor performed the best with an explained variance score of 0.778, an R^2 of 0.770, an RMSE of 92.95 minutes, and an MAE of 44.31 minutes. Based on SHAP analysis of the XGBoost regression, body mass index, spinal fusions, surgical procedure, and number of spine levels involved were the features with the most impact on the model.

Conclusions: Using ensemble learning-based predictive models, specifically XGBoost regression, can improve the accuracy of the estimation of spine surgery times.

(JMIR Perioper Med 2023;6:e39650) doi:10.2196/39650

XSL•FO
RenderX

KEYWORDS

ensemble learning; machine learning; spine surgery; case duration; prediction accuracy; operating room efficiency; learning; surgery; spine; operating room; case; model; patient; surgeon; linear regression; accuracy; estimation; time

Introduction

Surgery is an important component of care for many patients experiencing pathology of the spine. Lower back pain, degenerative disease of the spine, and other related ailments cost the United States tens of billions of dollars a year in direct medical expenses and lost productivity [1]. Martin and colleagues [2] reported the incidence of elective fusion of the lumbar spine increasing over 60% from 2004 to 2015, with hospital costs for such surgeries surging over 170% in the same time to an average of over US $50,000 per admission. Despite new trends in cost containment [3-5], new operative techniques, expansion of surgical navigation and imaging systems, implementation of specialized postoperative recovery pathways, and increased demand for services in an aging patient population have resulted in a complex, highly variable operational environment [6-9]. Such heterogeneity can make planning and use of resources challenging. The operating room is a critical target for decreasing costs and is second only to the patient room and board in the total expense of a perioperative episode [10]. Many strategies for improving operating room efficiency focus on time management [11,12]. Predicted surgical case duration often informs how cases are scheduled and which resources are dedicated to prepare for and staff them [13]. Consequently, improving the accuracy of these predictions is a practical strategy to increase operating room efficiency [14].

Surgeons often estimate case durations when scheduling operative time; durations may also be tied to historical averages or Current Procedural Terminology (CPT) codes, practices that are prone to substantial inaccuracies [15]. Classical statistical methods have been used to further improve the prediction of case durations [16-18]. The proliferation of electronic health records and the associated generation of vast amounts of previously uncaptured patient data have allowed for more sophisticated analytics in several clinical arenas, including the operating room [19]. With large enough data sets, specialized algorithms can develop complex predictive models after being exposed to a number of prior examples in a process known as machine learning [20].

Current predictive techniques in spine surgery include less sophisticated approaches such as classical multivariable statistical models. While a variety of features and outcomes, such as length of stay, prescription duration, and time to return to normal work, have been predicted in previous studies, there has been little focus on case duration [21-25]. To our knowledge, no other studies have focused on using machine learning models to predict the surgical case duration for the spine surgery population, but the method has been implemented in other procedures [26-28]. Spine surgery consists of heterogenous anatomical and technical components that should theoretically be taken into account when estimating case duration. The primary objective of this study is to develop machine learning-based predictive models using patient and surgery-specific features. Specifically, we use ensemble learning, which combines multiple predictive models to determine an overall prediction of the outcome. We hypothesize that such models can outperform those that estimate case duration based on historic averages and surgeon preference (which may not be scalable or transferable outside of a given institution).

Methods

Ethics Approval

This retrospective study was approved (approval protocol 210098) by the Human Research Protections Program at the University of California, San Diego for the collection of data from our electronic medical record system. For this study, the informed consent requirement was waived. Data were collected retrospectively from the electronic medical record system of our institution's operating room data. Data from all patients that underwent spine surgery from 3 different orthopedic spine surgeons from January 2018 to September 2021 was extracted. We excluded all patients that had missing data for actual case duration; all other features with missing values were categorized as unknown or imputed if they were continuous variables (described below). This retrospective observational study abided by the EQUATOR guidelines.

Primary Objective and Data Collection

The primary outcome measurement was a continuous value, defined as the actual operating room case duration measured in minutes (from patient wheeling into the operating room to exiting the operating room). We implemented predictive models using various machine learning algorithms to predict the actual case duration. We compared this to our current system's practice of estimating case duration, which is equal to the mean of the last 3 times the surgical procedure was performed, with the ability of the surgeon to change times based on their preference. The models developed were multivariable linear regression, random forest regressors, bagging regressors, and XGBoost (Extreme Gradient Boosting) regressors.

The independent features in the models were (1) categorical features, which included surgical procedure (39 unique procedures), surgeon identification (3 different surgeons), American Society of Anesthesiologists Physical Status (ASA PS) score (ie, comorbidity burden), sex, specific surgical details (kyphoplasty, discectomy, fusion, and laminectomy), the anterior approach involved (ie, approach surgeon used to access the spine), and level of spine region involved (eg, cervical, thoracic, lumbar, or a combination of levels); and (2) continuous features, which included the number of spine levels involved in the surgery (from 1 to 7) and body mass index (kg/m^2) (Table 1). For missing data on the ASA PS class, the value was defined as "unknown." For missing data on body mass index, the value was imputed by using the average BMI among all patients with known data for this feature.

XSL•FO
RenderX

Table 1. Characteristics of all the cases included in analysis (N=3315).

Characteristics	Instrumentation	Approach	Fusion	Levels	Other	Participants, n (%)
Surgical Procedure						
Discectomy	No	Anterior	Yes	1	Fusion	89 (2.7)
Discectomy	No	Anterior	Yes	2	Fusion	127 (3.8)
Discectomy	No	Anterior	Yes	3+	Fusion	202 (6.1)
Deformity fusion	No	Posterior	Yes	1-6 seg.	For deformity	8 (0.2)
Deformity fusion	No	Posterior	Yes	7-12 seg.	For deformity	2 (0.1)
Lumbar fusion	No	Anterior	Yes	2	Lumbar	270 (8.1)
Lumbar fusion	No	Anterior	Yes	3	Lumbar	14 (0.4)
Oblique lumbar interbody fusion	No	Anterior	Yes	1	Lumbar	1 (0.0)
Transforaminal lumbar interbody fusion	No	Transforaminal	Yes	1	Lumbar	9 (0.3)
Extreme lateral interbody fusion	No	Lateral	Yes	1	Lumbar	251 (7.6)
Extreme lateral interbody fusion	No	Lateral	Yes	2	Lumbar	198 (6.0)
Extreme lateral interbody fusion	No	Lateral	Yes	3	Lumbar	63 (1.9)
Extreme lateral interbody fusion	No	Lateral	Yes	4	Lumbar	16 (0.5)
Thoracic fusion	No	Posterior	Yes	1	Thoracic	1 (0.0)
Thoracic fusion	No	Posterior	Yes	2	Thoracic	7 (0.2)
Thoracic fusion	No	Posterior	Yes	3	Thoracic	17 (0.5)
Thoracic fusion	No	Posterior	Yes	4	Thoracic	12 (0.4)
Thoracic fusion	No	Posterior	Yes	5+	Thoracic	35 (1.1)
Kyphoplasty or vertebroplasty	No	N/A	No	1	All	316 (9.5)
Kyphoplasty	No	N/A	No	2	Thoracolumbar	40 (1.2)
Kyphoplasty	No	N/A	No	3	Thoracolumbar	19 (0.6)
Kyphoplasty	No	N/A	No	4	Thoracolumbar	21 (0.6)
Laminectomy or decompressive laminectomy	No	Posterior	No	1	Lumbar	148 (4.5)
Laminectomy or decompressive laminectomy	No	Posterior	No	2	Lumbar	106 (3.2)
Laminectomy or decompressive laminectomy	No	Posterior	No	3	Lumbar	110 (3.3)
Laminectomy	No	Posterior	Yes	5	Cervical	109 (3.3)
Laminectomy	No	Posterior	Yes	1-4	Cervical	115 (3.5)
Laminectomy	No	Posterior	No	1-2	Cervical	3 (0.1)
Laminectomy	No	Posterior	No	2+	Cervical	31 (0.9)
Laminectomy	Yes	Posterior	Yes	1	Lumbar	219 (6.6)
Laminectomy	Yes	Posterior	Yes	2	Lumbar	259 (7.8)
Laminectomy	Yes	Posterior	Yes	3	Lumbar	162 (4.9)
Laminectomy	Yes		Yes	4+	Lumbar	259 (7.8)
Laminectomy	Yes	Posterior	Yes	1	Thoracic	9 (0.3)
Laminectomy	Yes	Posterior	Yes	2	Thoracic	9 (0.3)
Laminectomy	Yes	Posterior	Yes	3	Thoracic	7 (0.2)
Laminectomy	Yes	Posterior	Yes	4	Thoracic	18 (0.5)
Laminectomy	Yes	Posterior	Yes	5	Thoracic	30 (0.9)
Laminectomy	Yes	Posterior	Yes	6+	Thoracic	3 (0.1)

XSL•FO
RenderX

Characteristics	Instrumentation	Approach	Fusion	Levels	Other	Participants, n (%)
Specific surgical procedure included						
Kyphoplasty	N/A[a]	N/A	N/A	N/A	N/A	396 (11.9)
Discectomy	N/A	N/A	N/A	N/A	N/A	418 (12.6)
Fusion	N/A	N/A	N/A	N/A	N/A	2521 (76.0)
Laminectomy	N/A	N/A	N/A	N/A	N/A	1597 (48.2)
Anterior approach involved	N/A	N/A	N/A	N/A	N/A	702 (21.1)
Number of spine levels involved						
1	N/A	N/A	N/A	N/A	N/A	1043 (31.5)
2	N/A	N/A	N/A	N/A	N/A	1050 (31.7)
3	N/A	N/A	N/A	N/A	N/A	594 (17.9)
4	N/A	N/A	N/A	N/A	N/A	441 (13.3)
5	N/A	N/A	N/A	N/A	N/A	174 (5.2)
6	N/A	N/A	N/A	N/A	N/A	11 (0.3)
7	N/A	N/A	N/A	N/A	N/A	2 (0.1)
Surgeon performing procedure						
A	N/A	N/A	N/A	N/A	N/A	1676 (50.6)
B	N/A	N/A	N/A	N/A	N/A	191 (5.8)
C	N/A	N/A	N/A	N/A	N/A	1448 (43.7)
Level of spine involved						
Cervical	N/A	N/A	N/A	N/A	N/A	567(17.1)
Thoracic	N/A	N/A	N/A	N/A	N/A	228 (6.9)
Lumbar	N/A	N/A	N/A	N/A	N/A	2165 (65.3)
Male sex	N/A	N/A	N/A	N/A	N/A	1800 (54.3)
BMI (kg/m^2), mean (SD)	N/A	N/A	N/A	N/A	N/A	29.7 (6.3)
ASA PS[b] classification score						
1	N/A	N/A	N/A	N/A	N/A	46 (1.4)
2	N/A	N/A	N/A	N/A	N/A	1140 (34.4)
3	N/A	N/A	N/A	N/A	N/A	2008 (60.6)
4	N/A	N/A	N/A	N/A	N/A	112 (3.4)
Unknown	N/A	N/A	N/A	N/A	N/A	9 (0.3)

[a]N/A: not applicable.

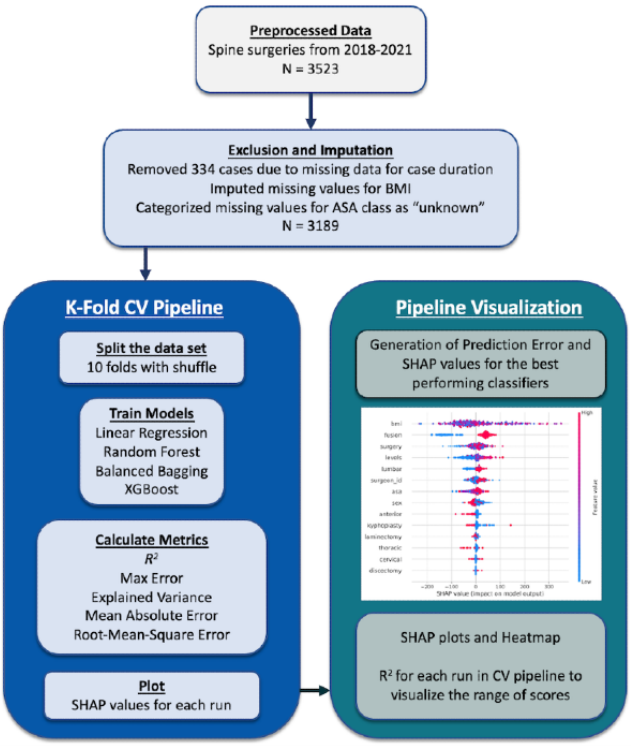
[b]ASA PS: American Society of Anesthesiologists Physical Status.

Analysis Packages and Metrics

Python (version 3.7.5; Python Software Foundation) was used for all statistical analyses. The code is provided in the webpage [29]. We calculated the R^2, root-mean-square error (RMSE), mean absolute error (MAE), explained variance, and maximum error for each iteration of k-fold cross-validation (described below) and used those scores to calculate the median scores and plot feature importance using SHAP (Shapley Additive Explanations) and prediction error plots (Figure 1).

XSL•FO
RenderX

Figure 1. Analysis pipeline-illustration of methodology. ASA: American Society of Anesthesiologists; CV: cross-validation; SHAP: Shapley Additive Explanations.



Machine Learning Models

Overview

We compared various machine learning-based predictive models to our institution's conventional model, which predicted case duration using average times (over the last 5 times the surgery was performed by that surgeon) based on the CPT code of the surgery plus adjustments from the surgical attending based on clinical judgment or preference. First, we developed a model using multivariable linear regression. We then evaluated the use of ensemble learning (a process in which multiple models are combined) to calculate a prediction. In this case, we used random forest, bagging, and XGBoost-based regressors (Multimedia Appendix 1). For each model, all features were included as inputs.

Multivariable Linear Regression

This is a statistical model that asserts a continuous outcome based on the weighted combination of the underlying independent variables. We tested an L2-penalty-based regression model without specifying individual class weights. This model provides a baseline score and helps make the case for improvement over the evaluation metrics.

Random Forest Regressor

Random forest is an ensemble approach (a technique that combines the predictions from multiple machine learning algorithms to make more accurate predictions than any individual model) of decision trees. It is a robust and reliable nonparametric supervised learning algorithm that acts as a means to test further improvements in metrics and determine the feature importance of a data set. The number of tree estimators was set to 1000, the criterion chosen was "squared error," and the minimum number of samples required to split an internal node was set to 2. All other parameters were left at their default values.

Bagging Regressor

Bagging or bootstrap aggregation is another way to build ensemble models. Bagging methods build several estimators on different randomly selected subsets of data. Unlike random forest models, bagging models are not sensitive to the specific data on which they are trained. They would give a similar score even when trained on a subset of the data. Bagging regressors are also generally more immune to overfitting. We built a bagging regressor using the scikit-learn package, where replacement was allowed. The number of estimators was set to 1000 with the base of decision tree regressors, and the samples were drawn with replacement (bootstrap was set to True). All other parameters were left at their default values.

XGBoost Regressor

Boosting is another approach to ensemble learning in which decision trees are built sequentially so that each subsequent tree aims to reduce the error from the previous tree. Thus, each subsequent tree learns from previous trees and updates the residual errors. Unlike bagging, boosting uses decision trees with fewer splits; XGBoost is an implementation of a gradient-boosted tree algorithm [30]. We built an XGBoost regressor using the *xgboost* version 1.7.1 package (xgboost developers). The number of estimators used was 1000, the tree method was set to "auto," and the booster was set to "gbtree." All other parameters were left at their default values, as described in the documentation of the library.

XSL•FO
RenderX

Feature Importance

An important function of a model is to uncover potential features that contribute to a given outcome. If a model can predict surgical outcomes efficiently with good specificity, then we can assume that the features of interest that are identified may be relevant and important to the actual surgical outcome. These models can often be opaque with many trees and features of interest, making interpretation of the data difficult. To aid in model interpretation, we used the SHAP model [31]. This module allows for a value to be assigned to each feature used to predict the outcome of a model. Additionally, it provides whether that feature negatively or positively impacts the outcome of that given prediction. If the score is very high or very low, that feature weighs heavily on the model. If the score is close to zero or not well separated, that feature is of lesser importance. Once features are identified and given SHAP values, interpretability is improved because features are concrete and have been assigned importance. Features can then be validated based on scientific rationale and further analysis.

k-Fold Cross-Validation

To perform a more robust evaluation of our models, we implemented k-fold cross-validation to observe the R^2, MAE, RMSE, explained variance, and maximum error for 10 folds after a shuffle. The data set was first shuffled to account for any sorting and then split into 10 folds, where 1 fold serves as the test set and the remaining 9 sets serve as the training set. This was repeated until all folds had the opportunity to serve as the test set. For each iteration, our performance metrics were calculated on the test set. The median of each performance metric (R^2, RMSE, MAE, explained variance, and maximum error) was calculated thereafter.

Results

Overview

There were 3523 spine surgeries identified during this period. After exclusion criteria were applied, 3189 surgeries were included in the final analysis. Among these, there were 39 different kinds of spine surgeries included. The majority of cases involved spinal fusion (n=2433, 76.0%) and were performed in the lumbar region (n=2082, 65.3%). The median ASA PS score was 3, and the majority of patients were male (n=1732, 54.3%; Table 1). The mean of actual surgical case duration among all surgeries was 335.5 (SD 199.9) minutes.

Performance Evaluation Using Linear Regression

Using all features (Table 1), we developed various machine learning algorithms to predict case duration. The base model, which was the conventional approach against which all machine learning models were compared, was based on our current system's method to predict surgical times, which is based on the average of the surgical procedures' case times over the last 5 instances with the ability for the surgeon to change times based on clinical judgment or preference. There was a poor coefficient of determination between the predicted time and actual time based on this approach (R^2=–0.213). We then performed multivariable linear regression trained on 80% of the data and tested on 20% of separate data, which had an R^2 of 0.34. Features that were statistically significant in this model included laminectomy (estimate=218.51, P<.001), number of levels performed, ASA PS classification score, and lumbar involvement (estimate=218.51, P<.001; Table 2).

Next, we implemented ensemble learning approaches to predicting case duration, in which the models were trained on 80% of the data and tested on a separate 20% of the data. The reason for the 80:20 split was to visualize the R^2 metric for each model (Figure 2). The R^2 metrics for the linear regressor, bagging regressor, random forest regressor, and XGBoost regressor, as well as the currently used method, were 0.407, 0.812, 0.812, 0.832, and 0.213, respectively.

XSL•FO
RenderX

Table 2. Results of the multivariable linear regression model predicting actual case duration. We included all features in the model. Because surgical procedure had 39 different procedures, we omitted the values from the table, however, they were included in the model.

	Estimate	SE (minutes)	*P* value
Intercept	–61.89	119.13	.60
Specific surgical procedure included			
Kyphoplasty	–225.43	90.5	.01
Discectomy	–33.94	94.7	.72
Fusion	6.17	75.9	.94
Laminectomy	218.51	41.7	<.001
Anterior approach involved	102.06	93.6	.28
Number of spine levels involved			
1	Reference		
2	62.57	77.8	.42
3	–18.15	80.6	.82
4	74.41	65.3	.25
5	246.18	61.7	<.001
6	212.36	106.7	.04
7	445.13	143.1	.002
Surgeon performing procedure			
A	Reference		
B	–38.37	13.1	.003
C	4.32	6.0	.47
Level of spine involved			
Cervical	115.84	53.6	.03
Thoracic	26.56	34.6	.44
Lumbar	218.51	52.3	<.001
Male sex	173.34	171.7	.31
BMI (kg/m^2)	0.38	0.52	.47
ASA PS[a] classification score			
1	Reference		
2	66.03	30.1	.32
3	97.43	25.9	<.001
4	29.98	30.1	.32
Unknown	41.97	60.2	.49

[a]ASA PS: American Society of Anesthesiologists Physical Status.

Figure 2. Illustration of the correlation between actual times and predicted surgical times for spine surgery calculated by each model type: predicted times based on procedural averages and surgeon preference or customization, multivariable linear regression, random forest, bagging, and Extreme Gradient Boosting (XGBoost). The data set was split 80:20 (training:test), and the model was trained on the training set and validated on the test set.

XSL•FO
RenderX

Median Performance Metrics of Models Using k-Fold Cross-Validation

We calculated various performance metrics for each model by applying a k-fold cross-validation approach and calculated the median scores for each model (Table 3). The linear regression model had an explained variance score of 0.34, an R^2 of 0.40, an RMSE of 162.84 minutes, and an MAE of 127.22 minutes. Among all models, the XGBoost regressor performed the best with an explained variance score of 0.778, an R^2 of 0.77, an RMSE of 92.95 minutes, and an MAE of 44.31 minutes.

SHAP analysis was performed to describe the features of the XGBoost model with the most impact on model prediction since it was the best-performing model based on the R^2 (Figure 3). Figure 3A illustrates the most important features per fold, whereas Figure 3B illustrates the ranks of each feature's importance per fold. BMI and spine fusion were consistently the top 2 most impactful features. In order of feature importance, there were then surgical procedure, number of spine levels, operating surgeon, the anatomic location being the lumbar spine, ASA PS classification score, sex, kyphoplasty, the anatomic location being the cervical spine, anterior approach, laminectomy, the anatomic location being the thoracic spine, and discectomy.

Table 3. Performance of each machine learning approach predicting case duration of spine surgery. Calculation is based on the median quantified by k-fold cross-validation for the bagging regressor, linear regression, random forest regressor, and XGBoost regressor. Current method is based on average of the last 5 instances of the surgery with surgeons input to modify time.

Model or method	Explained variance	Max error	MAE[a] (minutes)	RMSE[b] (minutes)	R^2
Current method	0.012	847	180.32	243.30	–0.57
Linear regression	0.345	526.29	127.21	162.84	0.34
RF[c] regressor	0.768	454.59	62.82	96.51	0.76
Bagging regressor	0.769	454.90	62.83	96.51	0.76
XGBoost regressor	0.778	475.72	44.31	92.95	0.77

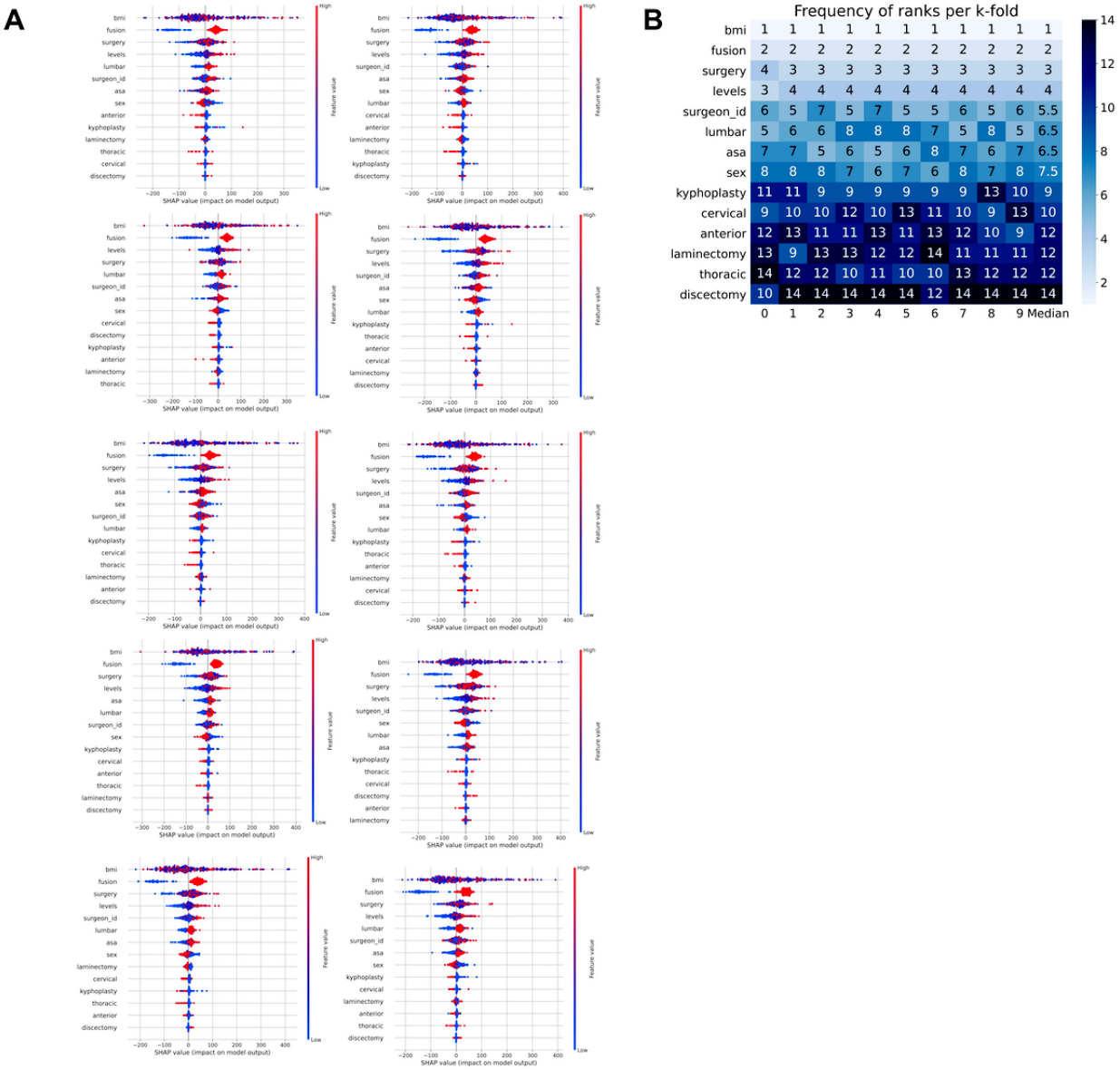
[a]MAE: mean absolute error.

[b]RMSE: root mean square error.

[c]RF: random forest.

XSL•FO
RenderX

Figure 3. Feature importance from the Extreme Gradient Boosting based on SHAP (Shapley Additive Explanations) values. (A) SHAP analysis for each of the 10 folds; (B) a heat map of the frequency of ranks for each feature per k-fold.



Discussion

Principal Findings

We found that the use of ensemble learning with the patient and procedural-specific features (variables that are known preoperatively and attainable from the electronic medical record system) outperformed the prediction of spine surgery case duration when compared to models that use historic averages and surgeon preferences. Unique to our approach of predicting surgical time for this heterogenous surgical population was the granularity of features (eg, patient and surgical characteristics) combined with an ensemble learning approach. The reference model (the time estimated based on historic averages and surgeon preference) had poor performance. We then implemented machine learning-based models using features including procedural details (ie, number of spine levels, patient positioning, surgeon, level of spine region involved, etc) and patient-specific details (ie, body mass index, sex, ASA score, etc) and demonstrated improved performance. While linear regression improved R^2 to 0.34, the use of XGBoost, random forest, and bagging improved it further (0.77, 0.71, and 0.71, respectively). Such models could be relatively easy to integrate into a resource-capable electronic medical record system, given that the included features could be obtained automatically from the electronic record preoperatively.

The usage of historic averages or CPT code-based estimations for spine surgery scheduling may be inaccurate given that some determinants of case duration may not be accounted for in the prediction. These features include surgeon experience, level of the spine region involved, number of levels, type of surgery (ie, kyphoplasty, fusion, laminectomy, etc), need for multiple surgeries, patient positioning, and patient body mass index. The inclusion of these features into our models results in a substantial improvement in prediction accuracy. Accurate prediction of operation times has long been discussed as a means to improve operating room efficiency and patient care [14]. Recent implementations of such models have demonstrated these improvements across a variety of measures. A recent randomized

XSL•FO
RenderX

clinical trial found that a machine-learning approach increased prediction accuracy, decreased start-time delay, and decreased total patient wait time [32]. A similar randomized controlled trial demonstrated increased throughput and decreased staff burnout [32,33]. Subsequently, decreases in delays and wait times result in lower costs and increased caseloads, which can further drive cost-effectiveness [34,35]. Associations between wait times and postoperative complications provide evidence that proper identification and mitigation of delays can improve outcomes as well [36]. Overall, improvements in patient scheduling, case duration, and staffing may result in enhanced efficiency and potentially superior patient outcomes. Understanding and identifying the features that are key in lessening the burden of misused surgical time is crucial with the trending increases in caseload burden and impacted hospital resources.

Ensemble learning essentially uses an "ensemble" of predictive models and calculates the overall prediction based on the individual predictions from each model within the "ensemble." In this case, we leveraged ensemble learning using decision tree-based machine learning algorithms: random forest, bagging, and boosting. Our results demonstrated a substantial improvement with XGBoost compared to the other ensemble approaches as well as linear regression. XGBoost often performs better than random forest because it prunes nodes if the gain of a node is minimal to the model [30]. Random forest generates the tree to a greater depth because it does not prune nodes and relies on a majority vote for the outcome. This can result in overfitting in random forest models. Random forest may also give preference to classes that are associated with categorical variables, which do not occur in XGBoost. Because XGBoost is an iterative process, it gives preference to features that enable the regressor to predict low-participation classes. Additionally, XGBoost is more efficient with the unbalanced data sets often seen in medical or biological data. Alternatives such as linear regression work well when the data is straightforward and well-distributed. The more complex the data set, the better a bagging or tree-based model will work. With ensemble approaches, nonlinear relationships between features may be captured, and a "strong" model is developed based on learning from "weaker" models, in which residual errors are improved. Thus, the use of ensemble learning in this clinical scenario—where there is a complex interplay between features—may be superior to a statistical approach that only models linear relationships (ie, linear regression). Future studies may benefit from other approaches such as support vector machines, which could be implemented to focus on accuracy, or penalized regressors, which could provide increased interpretability.

Oftentimes, machine learning approaches are described as "black boxes" because the interpretation of the importance of features to the predictive model is challenging. The implementation of an explainer model such as SHAP values is one way to elucidate the importance of features. In this study, SHAP identified that BMI is the most important feature of the model and provides weight and context to the feature about the other identified features [37,38]. BMI may be associated with increased case duration due to the additional technical and positioning challenges. Sex was also identified as an important feature. This finding is congruent with current research that demonstrates women are more likely to have bone loss earlier than men, and bone loss has been shown to affect surgical outcomes and recovery due to poor bone remodeling and healing [39,40]. Other interesting features with an important impact included the operating surgeons themselves, the ASA PS classification score, and the number of spine levels operated. It makes sense to include surgeons as a feature in predictive modeling as each physician may have different styles and comfort levels that could impact surgical time. The ASA PS classification score represents a patient's comorbidity burden and could suggest that patients with a higher comorbidity burden would require longer anesthesia times. Finally, it makes sense that the number of spine levels contributes to case duration, as this has a potentially linear relationship to how long surgery would take. Being able to put various features into the context of the research question is essential for translating the findings into actionable metrics. Overall, the SHAP analysis identified clinically relevant features for future exploration and evaluation.

There are several limitations to the study, mainly its retrospective nature; thus, the collection and accuracy of the data are only as reliable as what is recorded in the electronic medical record system. The current institutional practice for estimating scheduled case duration was based on the historic averages of the last 5 surgeries, with the surgeon's ability to change the times based on clinical judgment or preference. We do not have data on why and when surgeons changed the times. In addition, there were some missing data for actual case duration, but this only led to the removal of 5.9% of the initial data set. There may also be several features not included in the models that may substantially contribute to time estimates, including surgical resident involvement (and their level of training) or surgical instruments used. Furthermore, there are other machine learning algorithms that we did not test, including support vector machines and penalized regressors. Despite these limitations, we were able to develop a predictive model using XGBoost with a high R^2 value (>0.7). These findings would need to be validated externally and prospectively to determine their generalizability to spine surgeries.

Conclusions

Operating room efficiency is a key factor in maintaining and growing institutional profits. Additionally, improvements in operating room efficiency contribute to enhanced patient care and satisfaction. Given the technical and anatomical heterogeneity in spine surgeries, it has been a challenge to predict case duration using conventional methods at our institution. This method can be applied in the future to standard and heterogenous surgical procedures with or without class imbalance to identify key obstacles to future surgical efficiency; however, it is crucial to develop robust models to more accurately predict schedule case length. In our study, we demonstrated that patient and surgical features that are easy to collect from the electronic medical record can improve the estimation of surgical times using machine learning-based predictive models. Future implementation of machine learning-based models presents an alternative pathway to use

XSL•FO
RenderX

electronic medical record data to advance surgical efficiency and enrich patient outcomes.

Acknowledgments

SS and OG are supported by the National Institutes of Health Grants DA043799 and DA044451. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Authors' Contributions

RAG, BH, and SS are responsible for the study design. RAG, BH, SS, ALD, JLT, and RW are responsible for the analysis and interpretation of data and drafting of the article. OG is responsible for supporting SS and for reviewing the article.

Conflicts of Interest

The University of California has received funding and product for other research projects from Epimed International (Farmers Branch, TX); Infutronics (Natick, MA); Precision Genetics (Greenville County, SC); and SPR Therapeutics (Cleveland, OH) for author RAG. The University of California San Diego is a consultant for Avanos (Alpharetta, GA), which RAG represents. SS is the founder of BrilliantBiome, Inc.

Multimedia Appendix 1
Illustration of how each ensemble learning algorithm computes the prediction: (A) random forest, (B) bagging, and (C) Extreme Gradient Boosting (XGBoost).
[PDF File (Adobe PDF File), 2634 KB - periop_v6i1e39650_app1.pdf]

References

1. Dagenais S, Caro J, Haldeman S. A systematic review of low back pain cost of illness studies in the United States and internationally. Spine J 2008;8(1):8-20. [doi: 10.1016/j.spinee.2007.10.005] [Medline: 18164449]
2. Martin BI, Mirza SK, Spina N, Spiker WR, Lawrence B, Brodke DS. Trends in lumbar fusion procedure rates and associated hospital costs for degenerative spinal diseases in the United States, 2004 to 2015. Spine 2019;44(5):369-376. [doi: 10.1097/BRS.0000000000002822] [Medline: 30074971]
3. Ugiliweneza B, Kong M, Nosova K, Huang KT, Babu R, Lad SP, et al. Spinal surgery: variations in health care costs and implications for episode-based bundled payments. Spine 2014;39(15):1235-1242. [doi: 10.1097/BRS.0000000000000378] [Medline: 24831503]
4. Dietz N, Sharma M, Alhourani A, Ugiliweneza B, Wang D, Nuño MA, et al. Bundled payment models in spine surgery: current challenges and opportunities, a systematic review. World Neurosurg 2019;123:177-183. [doi: 10.1016/j.wneu.2018.12.001] [Medline: 30553071]
5. Mok J, Martinez M, Smith H, Sciubba D, Passias P, Schoenfeld A, Association for Collaborative Spine Research Investigators. Impact of a bundled payment system on resource utilization during spine surgery. Int J Spine Surg 2016;10:19 [FREE Full text] [doi: 10.14444/3019] [Medline: 27441177]
6. Dagal A, Bellabarba C, Bransford R, Zhang F, Chesnut RM, O'Keefe GE, et al. Enhanced perioperative care for major spine surgery. Spine 2019;44(13):959-966. [doi: 10.1097/BRS.0000000000002968] [Medline: 31205177]
7. Lamperti M, Tufegdzic B, Avitsian R. Management of complex spine surgery. Curr Opin Anaesthesiol 2017;30(5):551-556. [doi: 10.1097/ACO.0000000000000494] [Medline: 28731875]
8. Basil GW, Wang MY. Trends in outpatient minimally invasive spine surgery. J Spine Surg 2019;5(Suppl 1):S108-S114 [FREE Full text] [doi: 10.21037/jss.2019.04.17] [Medline: 31380499]
9. O'Lynnger TM, Zuckerman SL, Morone PJ, Dewan MC, Vasquez-Castellanos RA, Cheng JS. Trends for spine surgery for the elderly: implications for access to healthcare in North America. Neurosurgery 2015;77(suppl 4):S136-S141. [doi: 10.1227/NEU.0000000000000945] [Medline: 26378351]
10. Childers CP, Maggard-Gibbons M. Understanding costs of care in the operating room. JAMA Surg 2018;153(4):e176233 [FREE Full text] [doi: 10.1001/jamasurg.2017.6233] [Medline: 29490366]
11. Fong AJ, Smith M, Langerman A. Efficiency improvement in the operating room. J Surg Res 2016;204(2):371-383. [doi: 10.1016/j.jss.2016.04.054] [Medline: 27565073]
12. Rothstein DH, Raval MV. Operating room efficiency. Semin Pediatr Surg 2018;27(2):79-85. [doi: 10.1053/j.sempedsurg.2018.02.004] [Medline: 29548356]
13. Overdyk FJ, Harvey SC, Fishman RL, Shippey F. Successful strategies for improving operating room efficiency at academic institutions. Anesth Analg 1998;86(4):896-906. [doi: 10.1097/00000539-199804000-00039] [Medline: 9539621]
14. Dexter F, Macario A. Applications of information systems to operating room scheduling. Anesthesiology 1996;85(6):1232-1234 [FREE Full text] [doi: 10.1097/00000542-199612000-00002] [Medline: 8968168]

XSL•FO
RenderX

15. Eijkemans MJC, van Houdenhoven M, Nguyen T, Boersma E, Steyerberg EW, Kazemier G. Predicting the unpredictable: a new prediction model for operating room times using individual characteristics and the surgeon's estimate. Anesthesiology 2010;112(1):41-49 [FREE Full text] [doi: 10.1097/ALN.0b013e3181c294c2] [Medline: 19952726]
16. Dexter F, Traub RD, Qian F. Comparison of statistical methods to predict the time to complete a series of surgical cases. J Clin Monit Comput 1999;15(1):45-51. [doi: 10.1023/a:1009999830753] [Medline: 12578061]
17. Kayis E, Wang H, Patel M, Gonzalez T, Jain S, Ramamurthi RJ, et al. Improving prediction of surgery duration using operational and temporal factors. AMIA Annu Symp Proc 2012;2012:456-462 [FREE Full text] [Medline: 23304316]
18. Wright IH, Kooperberg C, Bonar BA, Bashein G. Statistical modeling to predict elective surgery time. Comparison with a computer scheduling system and surgeon-provided estimates. Anesthesiology 1996;85(6):1235-1245 [FREE Full text] [doi: 10.1097/00000542-199612000-00003] [Medline: 8968169]
19. Pramanik MI, Lau RY, Azad MAK, Hossain MS, Chowdhury MKH, Karmaker BK. Healthcare informatics and analytics in big data. Expert Syst Appl 2020;152:113388. [doi: 10.1016/j.eswa.2020.113388]
20. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med 2019;380(14):1347-1358. [doi: 10.1056/NEJMra1814259] [Medline: 30943338]
21. Saravi B, Hassel F, Ülkümen S, Zink A, Shavlokhova V, Couillard-Despres S, et al. Artificial intelligence-driven prediction modeling and decision making in spine surgery using hybrid machine learning models. J Pers Med 2022;12(4):509 [FREE Full text] [doi: 10.3390/jpm12040509] [Medline: 35455625]
22. Gruskay JA, Fu M, Bohl DD, Webb ML, Grauer JN. Factors affecting length of stay after elective posterior lumbar spine surgery: a multivariate analysis. Spine J 2015;15(6):1188-1195. [doi: 10.1016/j.spinee.2013.10.022] [Medline: 24184639]
23. Lubelski D, Ehresman J, Feghali J, Tanenbaum J, Bydon A, Theodore N, et al. Prediction calculator for nonroutine discharge and length of stay after spine surgery. Spine J 2020;20(7):1154-1158. [doi: 10.1016/j.spinee.2020.02.022] [Medline: 32179154]
24. Karhade AV, Cha TD, Fogel HA, Hershman SH, Tobert DG, Schoenfeld AJ, et al. Predicting prolonged opioid prescriptions in opioid-naïve lumbar spine surgery patients. Spine J 2020;20(6):888-895. [doi: 10.1016/j.spinee.2019.12.019] [Medline: 31901553]
25. Singh S, Ailon T, McIntosh G, Dea N, Paquet J, Abraham E, et al. Time to return to work after elective lumbar spine surgery. J Neurosurg Spine 2021:1-9. [doi: 10.3171/2021.2.SPINE202051] [Medline: 34560636]
26. Hung AJ, Chen J, Che Z, Nilanon T, Jarc A, Titus M, et al. Utilizing machine learning and automated performance metrics to evaluate robot-assisted radical prostatectomy performance and predict outcomes. J Endourol 2018;32(5):438-444. [doi: 10.1089/end.2018.0035] [Medline: 29448809]
27. Zhao B, Waterman RS, Urman RD, Gabriel RA. A machine learning approach to predicting case duration for robot-assisted surgery. J Med Syst 2019;43(2):32. [doi: 10.1007/s10916-018-1151-y]
28. Tuwatananurak JP, Zadeh S, Xu X, Vacanti JA, Fulton WR, Ehrenfeld JM, et al. Machine learning can improve estimation of surgical case duration: a pilot study. J Med Syst 2019;43(3):44. [doi: 10.1007/s10916-019-1160-5] [Medline: 30656433]
29. bhavyahh/Spine_Surgery_Duration: Submitted for Manuscript Review. Zenodo. URL: https://doi.org/10.5281/zenodo.7369286 [accessed 2023-01-04]
30. Nwanosike EM, Conway BR, Merchant HA, Hasan SS. Potential applications and performance of machine learning techniques and algorithms in clinical practice: a systematic review. Int J Med Inform 2022;159:104679. [doi: 10.1016/j.ijmedinf.2021.104679] [Medline: 34990939]
31. Lundberg S, Lee SI. A unified approach to interpreting model predictions. arXiv. URL: http://arxiv.org/abs/1705.07874 [accessed 2023-12-01]
32. Strömblad CT, Baxter-King RG, Meisami A, Yee SJ, Levine MR, Ostrovsky A, et al. Effect of a predictive model on planned surgical duration accuracy, patient wait time, and use of presurgical resources: a randomized clinical trial. JAMA Surg 2021;156(4):315-321 [FREE Full text] [doi: 10.1001/jamasurg.2020.6361] [Medline: 33502448]
33. Kougias P, Tiwari V, Sharath SE, Garcia A, Pathak A, Chen M, et al. A statistical model-driven surgical case scheduling system improves multiple measures of operative suite efficiency: findings from a single-center, randomized controlled trial. Ann Surg 2019;270(6):1000-1004. [doi: 10.1097/sla.0000000000002763]
34. Aiken T, Barrett J, Stahl CC, Schwartz PB, Udani S, Acher AW, et al. Operative delay in adults with appendicitis: time is money. J Surg Res 2020;253:232-237. [doi: 10.1016/j.jss.2020.03.038] [Medline: 32387570]
35. Hopkins RB, Tarride JE, Bowen J, Blackhouse G, O'Reilly D, Campbell K, et al. Cost-effectiveness of reducing wait times for cataract surgery in Ontario. Can J Ophthalmol 2008;43(2):213-217. [doi: 10.3129/i08-002] [Medline: 18347625]
36. Pincus D, Ravi B, Wasserstein D, Huang A, Paterson JM, Nathens AB, et al. Association between wait time and 30-day mortality in adults undergoing hip fracture surgery. JAMA 2017;318(20):1994-2003 [FREE Full text] [doi: 10.1001/jama.2017.17606] [Medline: 29183076]
37. Azimi P, Yazdanian T, Shahzadi S, Benzel EC, Azhari S, Nayeb Aghaei H, et al. Cut-off value for body mass index in predicting surgical success in patients with lumbar spinal canal stenosis. Asian Spine J 2018;12(6):1085-1091 [FREE Full text] [doi: 10.31616/asj.2018.12.6.1085] [Medline: 30322247]
38. Jackson KL2, Devine JG. The effects of obesity on spine surgery: a systematic review of the literature. Global Spine J 2016;6(4):394-400 [FREE Full text] [doi: 10.1055/s-0035-1570750] [Medline: 27190743]

XSL•FO
RenderX

39. Alswat KA. Gender disparities in osteoporosis. J Clin Med Res 2017;9(5):382-387 [FREE Full text] [doi: 10.14740/jocmr2970w] [Medline: 28392857]
40. Park SB, Chung CK. Strategies of spinal fusion on osteoporotic spine. J Korean Neurosurg 2011;49(6):317-322. [doi: 10.1016/0008-8749(90)90007-e] [Medline: 2188738]

Abbreviations

ASA PS: American Society of Anesthesiologists Physical Status
CPT: Current Procedural Terminology
MAE: mean absolute error
RMSE: root-mean-square error
SHAP: Shapley Additive Explanations
XGBoost: Extreme Gradient Boosting

Edited by T Leung; submitted 16.05.22; peer-reviewed by A Das, M Michelson, Z Shahid, P Giabbanelli; comments to author 17.08.22; revised version received 29.11.22; accepted 25.12.22; published 26.01.23.

Please cite as:
Gabriel RA, Harjai B, Simpson S, Du AL, Tully JL, George O, Waterman R
An Ensemble Learning Approach to Improving Prediction of Case Duration for Spine Surgery: Algorithm Development and Validation
JMIR Perioper Med 2023;6:e39650
URL: https://periop.jmir.org/2023/1/e39650
doi:10.2196/39650
PMID:36701181

©Rodney Allanigue Gabriel, Bhavya Harjai, Sierra Simpson, Austin Liu Du, Jeffrey Logan Tully, Olivier George, Ruth Waterman. Originally published in JMIR Perioperative Medicine (http://periop.jmir.org), 26.01.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Perioperative Medicine, is properly cited. The complete bibliographic information, a link to the original publication on http://periop.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Assessing Barriers to Implementation of Machine Learning and Artificial Intelligence–Based Tools in Critical Care: Web-Based Survey Study

Eric Mlodzinski[1], MD; Gabriel Wardi[1,2], MD, MPH; Clare Viglione[3], MPH, RD; Shamim Nemati[4], PhD; Laura Crotty Alexander[1,5], MD; Atul Malhotra[1], MD

[1]Division of Pulmonary, Critical Care, Sleep and Physiology, University of California, San Diego, CA, United States

[2]Department of Emergency Medicine, University of California, San Diego, CA, United States

[3]Dissemination and Implementation Science Center, Altman Clinical & Translational Research Institute, University of California, San Diego, CA, United States

[4]Department of Biomedical Informatics, University of California, San Diego, CA, United States

[5]Section of Pulmonary and Critical Care, Veterans Affairs San Diego Healthcare System, San Diego, CA, United States

Corresponding Author:
Eric Mlodzinski, MD
Division of Pulmonary, Critical Care, Sleep and Physiology
University of California
9300 Campus Point Drive
San Diego, CA, 92037
United States
Phone: 1 858 657 7118
Email: emlodz@gmail.com

Abstract

Background: Although there is considerable interest in machine learning (ML) and artificial intelligence (AI) in critical care, the implementation of effective algorithms into practice has been limited.

Objective: We sought to understand physician perspectives of a novel intubation prediction tool. Further, we sought to understand health care provider and nonprovider perspectives on the use of ML in health care. We aim to use the data gathered to elucidate implementation barriers and determinants of this intubation prediction tool, as well as ML/AI-based algorithms in critical care and health care in general.

Methods: We developed 2 anonymous surveys in Qualtrics, 1 single-center survey distributed to 99 critical care physicians via email, and 1 social media survey distributed via Facebook and Twitter with branching logic to tailor questions for providers and nonproviders. The surveys included a mixture of categorical, Likert scale, and free-text items. Likert scale means with SD were reported from 1 to 5. We used student t tests to examine the differences between groups. In addition, Likert scale responses were converted into 3 categories, and percentage values were reported in order to demonstrate the distribution of responses. Qualitative free-text responses were reviewed by a member of the study team to determine validity, and content analysis was performed to determine common themes in responses.

Results: Out of 99 critical care physicians, 47 (48%) completed the single-center survey. Perceived knowledge of ML was low with a mean Likert score of 2.4 out of 5 (SD 0.96), with 7.5% of respondents rating their knowledge as a 4 or 5. The willingness to use the ML-based algorithm was 3.32 out of 5 (SD 0.95), with 75% of respondents answering 3 out of 5. The social media survey had 770 total responses with 605 (79%) providers and 165 (21%) nonproviders. We found no difference in providers' perceived knowledge based on level of experience in either survey. We found that nonproviders had significantly less perceived knowledge of ML (mean 3.04 out of 5, SD 1.53 vs mean 3.43, SD 0.941; $P<.001$) and comfort with ML (mean 3.28 out of 5, SD 1.02 vs mean 3.53, SD 0.935; $P=.004$) than providers. Free-text responses revealed multiple shared concerns, including accuracy/reliability, data bias, patient safety, and privacy/security risks.

Conclusions: These data suggest that providers and nonproviders have positive perceptions of ML-based tools, and that a tool to predict the need for intubation would be of interest to critical care providers. There were many shared concerns about ML/AI in health care elucidated by the surveys. These results provide a baseline evaluation of implementation barriers and determinants

XSL•FO
RenderX

of ML/AI-based tools that will be important in their optimal implementation and adoption in the critical care setting and health care in general.

(JMIR Perioper Med 2023;6:e41056) doi:10.2196/41056

KEYWORDS

surveys and questionnaires; machine learning; artificial intelligence; critical care; respiratory insufficiency; survey; Qualtrics; questionnaire; perception; trust; perspective; attitude; intubation; predict; barrier; adoption; implementation

Introduction

Machine learning (ML) is increasingly used for the development of predictive models in health care, although implementation into clinical care has been limited [1-5]. We have recently reported a deep learning algorithm to predict the need for intubation in patients at risk of respiratory failure in the intensive care unit (ICU) [6]. This algorithm was validated on multiple data sets and was shown to outperform expert clinicians as well as an established predictive model [7]. However, this algorithm is not yet widely implemented.

ML algorithms have been published across nearly all fields of medicine, with models developed for the interpretation of clinical imaging and pathology slides, to assist in the diagnosis of skin lesions, and to predict clinical decompensation and mortality risks in specific populations [8]. In the pulmonary and critical care space, there have been prediction models developed to identify and risk stratify pulmonary nodules on computed tomography scans, and sepsis prediction algorithms to detect clinical decompensation prior to patients meeting clinical sepsis criteria [9-12]. Mechanical ventilation is another area that has seen a growing number of algorithms, with particular focus on predicting successful weaning, ventilator-associated complications, ventilator asynchrony, and timing of need for intubation [3]. However, despite the novelty and potential utility of these models, most have not been used in patient care [1,2]. Thus, further efforts to elucidate barriers and facilitators to implementation are clearly warranted.

There have been a small number of studies evaluating perceptions of ML-based tools among providers or nonproviders using surveys or semistructured interviews. One study of 12 providers in the United Kingdom assessing perception of artificial intelligence (AI) in the National Health Services found concerns over a lack of infrastructure, funding, and a common language [4]. Another study of general practitioners in the United Kingdom suggested that these clinicians felt there was only limited potential for these technologies in their practice [13]. Richardson et al [14] assessed patient perceptions of AI in health care via semistructured interviews and found an overall positive perception of advances in health care technology but also concerns over safety, costs, and patient autonomy. Another study evaluating patient perceptions of ML/AI being used in a skin cancer screening tool found a favorable reception of this technology but only if being used to assist and not replace physician judgment [15]. A similar analysis of patient perception of ML/AI in skeletal radiography revealed a strong preference toward physician interpretation over an AI tool [16]. To our knowledge, there are no studies assessing potential implementation determinants of such tools in the critical care setting.

To identify barriers and facilitators to implementation of a novel tool that predicts the need for mechanical ventilation, as well as to better understand perceptions of ML-based tools across health care, we emailed surveys to providers and shared surveys via social media for both nonproviders and providers. We gathered qualitative information from the surveys to identify potential barriers, which may need to be addressed prior to optimal implementation of these approaches. We further sought to determine whether providers' level of confidence with and perceived knowledge of ML would be a function of their level of experience. Overall, based on our experience and prior studies, we hypothesized that senior physicians may be less comfortable with ML algorithms compared to their more junior counterparts [16]. Second, we hypothesized that nonproviders would be more skeptical of machine learning tools than the providers who may be using them.

Methods

Survey Development

Our survey items were created and reviewed by a team of 4 critical care providers, 1 machine learning expert, and 1 implementation science expert to ensure completeness, functionality, and appropriate format based on published recommendations for surveys [17-21]. Survey structure and questions were not altered after survey dissemination. Respondents were provided with informed consent (Multimedia Appendices 1 and 2) and had the option to remain completely anonymous.

Ethical Considerations

The University of California, San Diego (UCSD) institutional review board reviewed the study and waived the need for approval (UCSD IRB Project #210349XX, "Survey of ICU Clinicians Regarding the Implementation of a Novel EMR-Based Algorithm to Predict Need for Mechanical Ventilation in ICU Patients," with an amendment for expanded survey with social media recruitment, initial waiver of approval date March 30, 2021, amendment waiver of approval date August 12, 2021).

Single-Center Critical Care Physician Survey

Our single-center physician survey (Multimedia Appendix 1) was an open, voluntary, anonymous questionnaire that consisted of 8 items and was distributed to 99 critical care physician trainees and faculty at our institution via email. The survey consisted of 3 pages of content, with 6 multiple-choice and 2 free-response questions. Likert scales of 1-5 were used for

XSL•FO
RenderX

opinion-based questions, with 1 representing the most negative and 5 the most positive outcome, and 3 representing a "moderate" response. The results are presented as means with SD. Likert scales were also converted to 3 groups with 1 and 2 representing "low," 3 representing "moderate," and 4 and 5 representing "high," and the percentage of each category was reported. Respondents could go back and change answers prior to submitting the survey if desired. Data were collected over a 2-week period in May 2021.

Social Media Survey

Our social media survey (Multimedia Appendix 2) was an open, voluntary, anonymous survey distributed via Twitter and Facebook posts (Multimedia Appendix 3) by our research team. The survey contained 3 pages of content and consisted of an initial question distinguishing medical providers from nonproviders, which then branched into an 11-question survey for providers and 10-question survey for nonproviders. Professions that were under the category of providers included: physicians (practicing or in-training), advanced practice providers, nurses, and medical students. Each survey included a mixture of multiple-choice and free-response questions. Likert scales of 1-5 were used for opinion-based questions, with 1 representing the most negative and 5 the most positive outcome, with 3 representing either a "neutral" or "moderate" response depending on the question. Outcomes are presented as means (SD). Likert scales were also converted to three groups with 1 and 2 representing "low" or "negative," 3 representing "moderate" or "neutral," and 4 and 5 representing "high" or "positive," and the percentage of each category was reported. One adaptive question was used. Respondents could go back and change answers prior to submitting the survey if desired. Providers were offered the chance to complete the survey as a nonprovider as well, although this was not tracked. Data were collected over a 1-month period from September to October 2021. An incentive of an Amazon gift card was offered in the social media survey for one of the respondents chosen randomly.

Survey Analyses

We used the CHERRIES (Checklist for Reporting Results of Internet E-Surveys) to guide our survey reporting and analysis (Multimedia Appendix 4) [22]. No view rate or participation rate was known for either survey. Data were collected and stored securely in Qualtrics for both surveys. The estimated time of each survey was approximately 5 minutes.

We performed a completeness check of our survey data and removed survey responses with <25% completion rate or response times of <30 seconds. No cookies were used in tracking responses. For our social media–based survey, we performed a quality analysis of the survey data and removed responses deemed suspicious for "bot" activity. We screened for suspicious responses by flagging responses with exact matching free-text responses with timestamps within a 4-hour period, and these responses were removed. We also screened for duplicate email addresses, and responses sharing the same email addresses were removed.

Secondary analyses were completed to determine if perceived knowledge of ML/AI and comfort with using ML/AI tools differed by level of provider experience or between nonproviders and providers. We also studied whether prior understanding of ML had an association with potential barriers to implementation of ML algorithms into clinical practice.

Data were analyzed using Excel (v.18.2110.13110.0; Microsoft Corporation) and SPSS Statistics (version 28; IBM Corp). Descriptive statistics were summarized as previously indicated. Independent *t* tests were used for comparison of means of each group of interest, with *t* statistic value, *df*, and the *P* value reported for each outcome. Cohen d_s (for unequal group sizes) values were calculated to estimate effect size between groups where significant differences existed. To minimize extreme responding bias and allow for binary analysis, certain Likert scale results were converted into binary format (using a score of 4-5 as a "positive" response and 1-3 as a "negative" response), which was based on similar methodology used in a previous survey-based study design [23]. Chi-square tests were used for comparing binary responses. Odds ratios with 95% confidence intervals are presented when applicable. A 2-sided $\alpha<.05$ was considered significant.

Qualitative Analysis

The free-text responses were reviewed by a member of the study team. We removed responses that were deemed to be uninterpretable (due to content unrelated to the topic or nonsensical language). The percentage of respondents who provided valid responses for our qualitative questions were determined. We performed content analysis for each free-text response for both surveys regarding concerns of use of ML/AI in practice and determined shared themes across all surveys. Each response could be categorized into one or more themes. Only themes with at least 5% of responses fitting within that category were reported.

Results

Single-Center Critical Care Physician Survey

Out of 79 physicians, 47 completed this internal institutional survey. The results of the survey are displayed in Tables 1 and 2 and Figure 1. All means are presented with SD. A total of 31 (59%) respondents were attendings, 19 (36%) were fellows, and 2 (4%) were residents. Perceived knowledge of ML was low (mean 2.40, SD 0.96), with 7.5% of respondents rating their knowledge as a 4 or 5. A total of 8 (15%) respondents had knowingly used an ML-based tool in their clinical practice. Confidence in predicting the need for mechanical ventilation due to COVID-19 pneumonia (mean 3.57, SD 0.79) was lower than for respiratory failure due to all other causes (mean 3.89, SD 0.78). Overall, willingness to use an ML-based algorithm was 3.32 (SD 0.95), with 75% of respondents answering 3 out of 5. Factors most likely to increase likelihood of utilization were "high quality evidence that it outperformed trained clinicians" (mean 4.28, SD 0.77), "transparency of the data utilized" (mean 4.13, SD 0.80), and "limited workflow interruption" (mean 4.09, SD 0.97), with more than 75% of respondents answering 4 or 5 for these 3 factors. For the free-response question 7 (regarding anticipated challenges with implementing an ML algorithm in practice), there were 18 (38%)

XSL•FO
RenderX

out of 47 valid responses, with 2 responses removed. Shared themes and responses per theme included: accuracy/reliability (n=7, 39%), workflow interruptions/alert fatigue (n=6, 33%), patient safety (n=1, 6%), and data bias (n=2, 11%). Representative examples are shown in Multimedia Appendix 5. For question 8 regarding suggestions on ways to improve implementation, 16 (34%) out of 47 participants provided valid responses, with no responses removed. Shared themes and responses per theme included: prospective data/proof of efficacy (n=10, 63%), electronic medical record integration (n=3, 19%), and data transparency (n=3, 19%).

Table 1. Characteristics of respondents to a single-center survey.

Characteristics	Respondents, n (%)
Response characteristics	
Response rate	53/99 (53)
Completion rate	47/53 (89)
Level of experience (n=53)	
Attending	31 (58)
Fellow	19 (36)
Resident	2 (3)
Other	1 (2)
Prior use of machine learning (n=53)	
Yes	8 (15)
No	35 (66)
Unsure	10 (18)

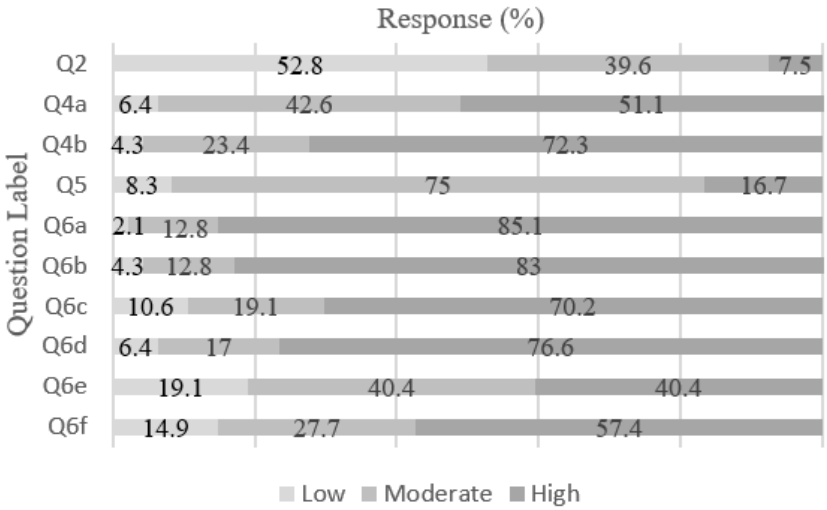
Table 2. Mean scores of Likert scale questions for a single-center survey.

Survey question	Score (1-5), mean (SD)
Q2. Level of knowledge of ML[a]	2.40 (0.968)
Q4a. Confidence in the ability to predict the need for mechanical ventilation in COVID-19	3.57 (0.801)
Q4b. Confidence in the ability to predict the need for mechanical ventilation for all other causes	3.89 (0.787)
Q5. Willingness to use ML-based tools to predict respiratory failure	3.32 (0.958)
Factors impacting the likelihood of using the tool	
Q6a. High-quality evidence available	4.28 (0.772)
Q6b. Limited workflow interruption	4.09 (0.974)
Q6c. Transparency of the data	4.13 (0.797)
Q6d. Real-time probability data of likelihood of need for mechanical ventilation	3.91 (0.974)
Q6e. Support from other intensive care unit clinicians and hospital leadership	3.57 (1.12)
Q6f. Standardized education on ML	3.34 (1.07)

[a]ML: machine learning.

XSL•FO
RenderX

Figure 1. Single-center survey Likert scale results. Responses were categorized into 3 separate categories (a response of 1 or 2 was considered "low," 3 "moderate," and 4 or 5 "high") and reported as percentage of valid responses out of 100%. Question content can be found in Table 2 and Multimedia Appendix 1.



Social Media Provider and Nonprovider Survey

We received 1196 responses, with 914 provider and 282 nonprovider responses. We excluded a total of 426 (35.6%) responses, 309 (33.8%) provider responses and 117 (41.5%) nonprovider responses. The reasons for exclusion included duplicate open-ended responses (n=324, 76.1%), duplicate email addresses (n=30, 7%), and <25% completion rate or <30-second time to completion (n=72, 16.9%). Our final analysis included a total of 770 total responses made up of 605 (78.6%) providers and 165 (21.4%) nonprovider responses. Descriptive results are displayed in Tables 3 and 4, Figures 2 and 3, and Multimedia Appendix 6. Physicians made up most of the respondents of the provider survey (n=372, 61.5%), with more attendings than medical trainees. A total of 21% (n=127) of respondents reported working in a critical care setting. Mean baseline understanding of ML/AI was 3.43 (SD 0.97), with 49% of respondents reporting a "high" level of knowledge. A total of 74% of respondents reported having used ML. Overall comfort with using an ML tool in patient care was 3.53 (SD 0.967), with 51.7% of respondents reporting a "high" level of comfort. Providers felt that efficiency and patient care were likely to benefit from these tools (means 3.68, SD 0.975 and 3.51, SD 0.938, respectively). Concern for potential negative impact on future jobs in medicine was 3.41 (SD 1.13), with 52% of respondents reporting "high" concern.

For the nonprovider survey, most nonproviders (n=90, 59%) had 1-5 encounters with the medical system in the last year. Overall confidence in physicians was 3.66 (SD 0.959), with 64.9% (n=107) reporting "high" confidence. Understanding of ML in health care was 3.03 (SD 1.23), with 45% (n=74) reporting understanding as "high" and 20.5% (n=34) as "low." Comfort with use of ML tools in health care was 3.27 (SD 1.01), with 57% (n=94) reporting "high" level of comfort. Nonproviders overall felt positively about how ML/AI would impact medical care (mean 3.40, SD 1.01), with 57% (n=94) having a "positive" response and 21.9% (n=36) having a "negative" response. The impact on relationship with their providers was reported "positive" in 35.8% (n=59) and "negative" in 27.2% (n=45) of respondents (mean 3.09, SD 0.931). A total of 74% (n=122) of respondents would want to know if an ML algorithm was being used in their care.

For the free-text question, providers and nonproviders were asked to share any concerns they had regarding ML or AI in health care. For the providers, 312 (52%) participants of 605 provided valid responses to the free-text question, with 28 responses removed. Of the 312 total responses, 56 (16.5%) reported no concerns and 256 (75.3%) responses included a concern. Shared themes and responses per theme included the following: accuracy/reliability (n=58, 22.7%), data bias (n=35, 13.7%), patient safety/outcomes (n=34, 13.3%), doctor-patient relationship (n=28, 10.9%), privacy/security (n=22, 8.6%), workflow (n=19, 7.4%), and costs (n=14, 5.5%). Representative examples are shown in Multimedia Appendix 5. For nonproviders, 109 (66%) participants provided valid free response, with 7 responses excluded. Of those 109 valid responses, 6 (5.5%) reported no concerns and 103 (94.5%) provided concerns. Shared themes and responses per theme included the following: accuracy/reliability (n=22, 21.4%), data bias (n=22, 21.4%), privacy/security (n=16, 15.5%), patient safety/outcomes (n=11, 10.7%), lack of knowledge of ML/AI (n=11, 10.7%), and doctor-patient relationship (n=10, 10.3%). Representative examples are shown in Multimedia Appendix 5.

XSL•FO
RenderX

Table 3. Likert scale responses of social media survey health care provider subgroup.

Survey question	Score (1-5), mean (SD)
Q2. How would you rate your current understanding of ML[a]/AI[b] as they apply to health care?	3.43 (0.970)
Q3a. How useful was this tool?	3.73 (0.917)
Q4. How comfortable would you feel using an ML or AI-based tool to make a clinical decision regarding your patients?	3.53 (0.967)
Q5. How concerned are you that ML/AI will make some health care jobs/specialties obsolete?	3.41 (1.13)
Please choose the option which best describes your opinion on how the implementation of ML/AI-based tools into routine clinical practice would impact each of the following	
Q6a. Patient care	3.51 (0.938)
Q6b. Efficiency in your daily practice	3.68 (0.975)
Q6c. Patient-provider relationship	3.34 (1.054)
Please rate the extent to which each of the following factors would increase your likelihood of using an ML or AI-based tool in your clinical practice	
Q7a. High-quality evidence of tool's efficacy	3.66 (0.995)
Q7b. Transparency of data	3.67 (1.02)
Q7c. Workflow interruptions	3.56 (1.06)
Q7d. Standardized education on ML/AI tools	3.63 (1.001)
Q7e. Support from administration	3.65 (1.025)

[a]ML: machine learning.

[b]AI: artificial intelligence.

Table 4. Likert scale responses of social media survey nonprovider subgroup.

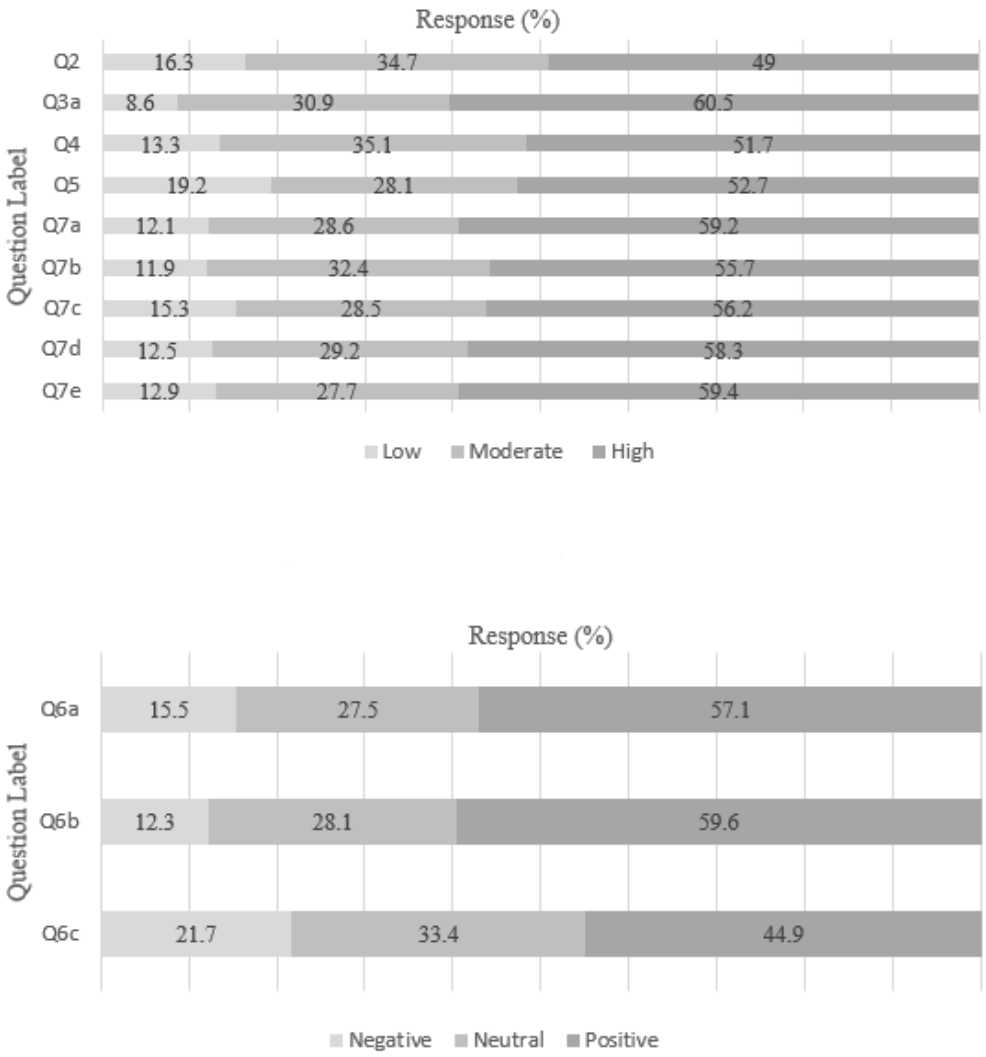
Survey question	Score (1-5), mean (SD)
Q2. How much confidence do you have in medical professionals' ability to make the correct decision for your medical care?	3.66 (0.959)
Q3. How would you rate your current understanding of ML[a] and AI[b] as they apply to health care?	3.03 (1.233)
Q4. How comfortable would you be with having a computer algorithm using ML/AI assisting in making decisions about your medical care?	3.27 (1.013)
Q5. How do you think the implementation of more ML/AI-based algorithms into the medical system will impact your medical care?	3.40 (1.014)
Q6. How do you think the implementation of more ML/AI-based algorithms into the medical system will impact your relationship with your medical team	3.09 (0.931)
Please rate the extent to which each of the following factors would increase your comfort level with an ML or AI-based tool being used in your medical care	
Q8a. High-quality evidence that it is as good or better than trained clinicians	3.65 (1.127)
Q8b. High-quality evidence that it can improve patient outcomes	3.77 (1.214)
Q8c. Knowing how the tool was developed	3.62 (1.082)
Q8d. Knowing that the tool would improve efficiency	3.56 (1.141)

[a]ML: machine learning.

[b]AI: artificial intelligence.

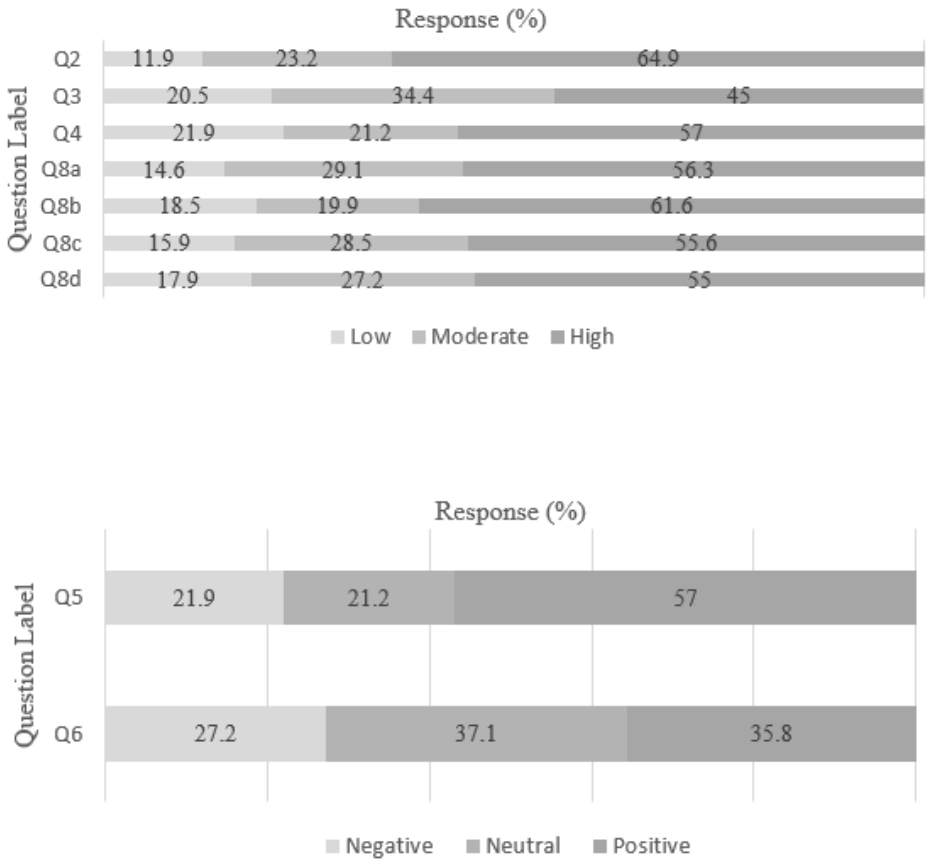
XSL•FO
RenderX

Figure 2. Provider survey Likert scale results. Responses were separated into 3 categories; "low," "moderate," or "high," depicted in the top graph, and "negative," "neutral," or "positive," depicted in the bottom graph. Results are reported as percentage of valid responses out of 100%. Question content can be found in Table 3 and Multimedia Appendix 2.



XSL•FO
RenderX

Figure 3. Nonprovider survey Likert scale results. Responses were separated into 3 categories; "low," "moderate," or "high," depicted in the top graph, and "negative," "neutral," or "positive," depicted in the bottom graph. Results are reported as percentage of valid responses out of 100%. Question content can be found in Table 4 and Multimedia Appendix 2.



Secondary Analyses

In the single-center survey, there was no significant difference between critical care trainees and attendings in terms of overall perceived knowledge of ML (mean 2.19, SD 0.991 vs mean 2.50, SD 0.862; t_{50}=1.19; *P*=.24) or willingness to use an ML prediction tool (mean 3.38, SD 1.05 vs 3.30, SD 0.907; t_{44}=0.248; *P*=.80). For the social media survey, there was no significant difference between trainees and attending physicians in perceived knowledge (mean 3.53, SD 0.840 vs mean 3.42, SD 1.03; t_{495}=1.20; *P*=.23) or comfort with ML tools (mean 3.60, SD 0.850 vs mean 3.51, SD 1.04; t_{492}=0.943; *P*=.35). There was a significant difference between physician and nonprovider knowledge of ML in health care (mean 3.43, SD 0.941 vs mean 3.04, SD 1.53; t_{752}=4.15; *P*<.001) and with comfort in using these tools (mean 3.53, SD 0.935 vs mean 3.28, SD 1.02; t_{746}=2.90; *P*=.004). Cohen d_s values were 0.33 and 0.28, respectively, suggesting a low effect size. Comparison of critical care physicians between the 2 surveys regarding their perceived knowledge of ML revealed a significantly lower perceived knowledge among the single-center survey respondents (mean 2.40, SD 0.936 vs mean 3.27 SD 1.01; t_{141}=5.08; *P*<.001). Cohen d_s value was 0.91, suggesting a large effect size. In a binary analysis of providers' baseline knowledge (high vs low), there was not a significant association between baseline knowledge and willingness to use ML in patient care (OR 2.270, 95% CI 0.694-7.424; *P*=.17). In a binary analysis of nonproviders' perceived knowledge of ML (high vs low), there was a significant association between higher knowledge of ML and more comfort with ML being used in patient care (OR 6.25, 95% CI 3.05-12.84; *P*<.001). The results are displayed in Table 5.

XSL•FO
RenderX

Table 5. Secondary analysis.

Secondary analysis subgroups	Score (1-5), mean (SD)	*t* score	*P* value[a]
Single-center survey			
Baseline knowledge of ML[b]		–1.19	.24
Attending	2.52 (0.991)		
Trainee	2.19 (0.862)		
Willingness to use ML tool		0.248	.81
Attending	3.30 (0.907)		
Trainee	3.38 (1.05)		
Social media survey			
Baseline knowledge of ML		–1.20	.23
Attending	3.42 (1.03)		
Trainee	3.53 (0.840)		
Comfort with using ML		0.943	.35
Attending	3.51 (1.04)		
Trainee	3.60 (0.850)		
Baseline knowledge of ML		4.15	<.001
Provider	3.43 (0.941)		
Nonprovider	3.04 (1.53)		
Comfort with using ML tool		2.90	.004
Provider	3.53 (0.935)		
Nonprovider	3.28 (1.02)		
Cross-survey analysis			
Baseline knowledge of ML		5.08	<.001
Critical care providers single center	2.4 (0.936)		
Critical care providers social media	3.27 (1.01)		
Chi-square analysis social media survey[c]			
Association of nonprovider comfort with using ML and baseline knowledge of ML	6.25 (3.05-12.84)	N/A[d]	<.001
Association of provider comfort with ML and baseline knowledge of ML	2.270 (0.694-7.424)	N/A	.167

[a]Independent student *t* tests were used for comparison of means; chi-square tests were used for comparing binary responses. $P<.05$ was considered significant.

[b]ML: machine learning.

[c]Odds ratios and 95% CI are provided for this category.

[d]N/A: not applicable.

Discussion

Principal Findings

To our knowledge, this is the first study to explore nonprovider and provider perspectives of novel ML-based tools in critical care as well as potential implementation determinants of these tools. We found that both providers and nonproviders have favorable attitudes toward the use of ML in health care, although there remained a small but significant difference between these 2 groups with providers having more comfort overall. Nonproviders with more perceived knowledge of the concept of ML/AI were more likely to feel favorable toward its use in patient care. This finding suggests that efforts to implement ML tools may require increased focus on nonprovider education and buy-in as skepticism may be more pronounced in this group [14,24]. Second, we observed no major difference in the level of knowledge or comfort among providers regardless of their level of experience in either survey, which contradicts our preconceived notions of older providers being less comfortable with technological advancements in medicine. Third, we identified nonprovider and provider concerns about potential systemic bias in data used in ML tools, patient safety, negative effects on the doctor-patient relationship, and data privacy/security. Among providers, we also identified workflow

XSL•FO
RenderX

interruptions as a major concern, and among nonproviders, limited knowledge of ML/AI was a major concern. These are critical factors that will need to be addressed to ensure user confidence in the data and algorithms. We also saw a large difference in comfort with ML among our own institution's critical care physicians compared to the more generalized critical care physicians, suggesting that institutional differences are likely to exist and that implementation methods may need to be tailored for each institution.

Our single-center survey provided important information regarding physician acceptance of a novel algorithm for predicting the onset of mechanical ventilation in patients at risk of respiratory failure. One of the goals of this survey was to serve as a needs assessment for this tool, and based on our results, it appears that providers at our institution feel that this tool would be beneficial in the clinical context of early prediction of the need for intubation among patients with respiratory failure due to COVID-19 and all other causes [6]. These results support our team's efforts in moving forward with the next steps of implementation, which will involve optimizing the interface for provider ease of use, preliminary prospective studies of its efficacy, and improving the sensitivity and specificity. Future steps include co-creating implementation strategies with a multidisciplinary team of clinicians, patients, implementation scientists, and medical informaticists to address identified determinants to improve the uptake and implementation of this algorithm. This process will also include additional surveys and structured interviews to assess ongoing effectiveness and to iteratively refine the algorithm to optimize its utility and improve clinical care.

Strengths and Limitations

One of the study's main strengths was the use of multiple platforms including social media for dissemination of our surveys, improving the generalizability of our results and allowing us to reach a large sample size. In addition, our surveys were unique in that we were able to gather both nonprovider and provider perspectives simultaneously. We also screened for suspicious responses in the social media surveys and removed these to increase the reliability of our survey results.

Despite our study's strengths, we acknowledge the following limitations. First, due to privacy issues, we did not collect demographic or other personal data regarding the respondents. Thus, we are unable to draw conclusions regarding whether certain members of nonprovider and provider communities may be more amenable to ML methods (eg, based on gender or race). Second, our conclusions are limited to the population studied as our surveys were in English and only reached those with electronic access. Third, as with any survey, there are risks of both selection bias as well as participation bias. For selection bias in the first survey, we emailed ICU providers but did not gather any systematic data from ICU nurses or pharmacists or others who may be impacted by these tools. Regarding participation bias, it is likely that the individuals responding to social media survey would be those with an interest in this topic and thus may be more comfortable with these methods than others. Fourth, the truly open nature of the social media survey led to unanticipated issues with bot responses, and while steps were taken to remove suspicious responses, to our knowledge, there is no validated means of screening for bot activity. Fifth, there was no specific implementation conceptual framework used in the development of questions addressing implementation barriers and facilitators. Despite these limitations, we view our findings as an important step toward the successful implementation of ML/AI methods to improve patient care.

Conclusions

Both providers and nonproviders have overall positive perspectives on the use of ML-based tools in health care, although nonproviders remain more skeptical. In addition, it appears that a tool to help predict onset of the need for intubation would be both useful and acceptable among critical care providers. Our study revealed shared concerns regarding accuracy and reliability, data bias, privacy/security, patient safety, the doctor-patient relationship, and workflow interruptions. These data provide a baseline assessment of health care provider and nonprovider perceptions of ML/AI-based tools that will be crucial in optimizing their clinical utility.

Acknowledgments

This study was supported by the National Institutes of Health (NIH) Research Project Grant (R01) HL154026 for which AM is the principal investigator. GW is funded by the NIH (R35GM143121 and K23GM146092) and the Gordon and Betty Moore Foundation (GBMF9052). SN is funded by the NIH (R56LM013517, R35GM143121, R01EB031539) and the Gordon and Betty Moore Foundation (GBMF9052). LCA is funded by the NIH (NHLBI K24 HL155884 and NHLBI R01HL147326) and a Veterans Affairs Merit Award (1I01BX004767). AM is funded by the NIH and reports income related to medical education from Livanova, Eli Lilly, and Jazz. ResMed, Inc provided a philanthropic donation to University of California, San Diego in support of a sleep center. EM and CV have no financial disclosures.

Data Availability

Data are stored securely in Qualtrics with access restricted to University of California, San Diego users. Email addresses were removed prior to exporting data to be processed in Excel and SPSS. Data can be requested by emailing the corresponding author at emlodzinski@health.ucsd.edu.

XSL•FO
RenderX

Authors' Contributions

All authors were involved in the conception and design of the study. EM was involved in data acquisition. EM, GW, and SN were involved in the data analysis. EM, GW, and AM were involved in manuscript writing. All authors were involved in manuscript review and final approval.

Conflicts of Interest

SN is a cofounder, advisor, and holds equity in Healcisio Inc, a startup company which is developing products related to the research described in this paper. The terms of this arrangement have been reviewed and approved by the University of California, San Diego in accordance with its conflict of interest policies.

Multimedia Appendix 1
Single-center critical care physician survey.
[DOCX File , 19 KB - periop_v6i1e41056_app1.docx]

Multimedia Appendix 2
Social media survey content.
[DOCX File , 22 KB - periop_v6i1e41056_app2.docx]

Multimedia Appendix 3
Social media recruitment post examples.
[DOCX File , 14 KB - periop_v6i1e41056_app3.docx]

Multimedia Appendix 4
Checklist for Reporting Results of Internet E-Surveys (CHERRIES).
[DOCX File , 19 KB - periop_v6i1e41056_app4.docx]

Multimedia Appendix 5
Qualitative evaluation of provider and patient concerns regarding ML/AI in healthcare.
[DOCX File , 23 KB - periop_v6i1e41056_app5.docx]

Multimedia Appendix 6
Demographics and categorical responses to social media survey.
[DOCX File , 25 KB - periop_v6i1e41056_app6.docx]

References

1. Seneviratne MG, Shah NH, Chu L. Bridging the implementation gap of machine learning in healthcare. BMJ Innov 2019;6(2):45-47. [doi: 10.1136/bmjinnov-2019-000359]
2. Verma AA, Murray J, Greiner R, Cohen JP, Shojania KG, Ghassemi M, et al. Implementing machine learning in medicine. CMAJ 2021;193(34):E1351-E1357 [FREE Full text] [doi: 10.1503/cmaj.202434] [Medline: 35213323]
3. Gallifant J, Zhang J, Del Pilar Arias Lopez M, Zhu T, Camporota L, Celi LA, et al. Artificial intelligence for mechanical ventilation: systematic review of design, reporting standards, and bias. Br J Anaesth 2022;128(2):343-351. [doi: 10.1016/j.bja.2021.09.025] [Medline: 34772497]
4. Morrison K. Artificial intelligence and the NHS: a qualitative exploration of the factors influencing adoption. Future Healthc J 2021;8(3):e648-e654 [FREE Full text] [doi: 10.7861/fhj.2020-0258] [Medline: 34888459]
5. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med 2019;380(14):1347-1358. [doi: 10.1056/NEJMra1814259] [Medline: 30943338]
6. Shashikumar SP, Wardi G, Paul P, Carlile M, Brenner LN, Hibbert KA, et al. Development and prospective validation of a deep learning algorithm for predicting need for mechanical ventilation. Chest 2021 Jun;159(6):2264-2273 [FREE Full text] [doi: 10.1016/j.chest.2020.12.009] [Medline: 33345948]
7. Roca O, Caralt B, Messika J, Samper M, Sztrymf B, Hernández G, et al. An index combining respiratory rate and oxygenation to predict outcome of nasal high-flow therapy. Am J Respir Crit Care Med 2019;199(11):1368-1376. [doi: 10.1164/rccm.201803-0589OC] [Medline: 30576221]
8. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019;25(1):44-56. [doi: 10.1038/s41591-018-0300-7] [Medline: 30617339]
9. Mlodzinski E, Stone DJ, Celi LA. Machine learning for pulmonary and critical care medicine: a narrative review. Pulm Ther 2020;6(1):67-77 [FREE Full text] [doi: 10.1007/s41030-020-00110-z] [Medline: 32048244]

XSL•FO
RenderX

10. Ardila D, Kiraly AP, Bharadwaj S, Choi B, Reicher JJ, Peng L, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nat Med 2019;25(6):954-961. [doi: 10.1038/s41591-019-0447-x] [Medline: 31110349]
11. Pehrson LM, Nielsen MB, Ammitzbøl Lauridsen C. Automatic pulmonary nodule detection applying deep learning or machine learning algorithms to the LIDC-IDRI database: a systematic review. Diagnostics (Basel) 2019;9(1):29 [FREE Full text] [doi: 10.3390/diagnostics9010029] [Medline: 30866425]
12. Nemati S, Holder A, Razmi F, Stanley MD, Clifford GD, Buchman TG. An interpretable machine learning model for accurate prediction of sepsis in the ICU. Crit Care Med 2018;46(4):547-553. [doi: 10.1097/CCM.0000000000002936] [Medline: 29286945]
13. Blease C, Kaptchuk TJ, Bernstein MH, Mandl KD, Halamka JD, DesRoches CM. Artificial intelligence and the future of primary care: exploratory qualitative study of UK general practitioners' views. J Med Internet Res 2019;21(3):e12802 [FREE Full text] [doi: 10.2196/12802] [Medline: 30892270]
14. Richardson JP, Smith C, Curtis S, Watson S, Zhu X, Barry B, et al. Patient apprehensions about the use of artificial intelligence in healthcare. NPJ Digit Med 2021;4(1):140 [FREE Full text] [doi: 10.1038/s41746-021-00509-1] [Medline: 34548621]
15. Nelson CA, Pérez-Chada LM, Creadore A, Li SJ, Lo K, Manjaly P, et al. Patient perspectives on the use of artificial intelligence for skin cancer screening: a qualitative study. JAMA Dermatol 2020;156(5):501-512. [doi: 10.1001/jamadermatol.2019.5014] [Medline: 32159733]
16. York T, Jenney H, Jones G. Clinician and computer: a study on patient perceptions of artificial intelligence in skeletal radiography. BMJ Health Care Inform 2020;27(3):e100233 [FREE Full text] [doi: 10.1136/bmjhci-2020-100233] [Medline: 33187956]
17. Karabenick SA, Woolley ME, Friedel JM, Ammon BV, Blazevski J, Bonney CR. Cognitive processing of self-report items in educational research: do they think what we mean? Educ Psychol 2007;42(3):139-151. [doi: 10.1080/00461520701416231]
18. Dennison CR, Mendez-Tellez PA, Wang W, Pronovost PJ, Needham DM. Barriers to low tidal volume ventilation in acute respiratory distress syndrome: survey development, validation, and results. Crit Care Med 2007;35(12):2747-2754. [doi: 10.1097/01.CCM.0000287591.09487.70] [Medline: 17901838]
19. Patel RP, Gambrell M, Speroff T, Scott TA, Pun BT, Okahashi J, et al. Delirium and sedation in the intensive care unit: survey of behaviors and attitudes of 1384 healthcare professionals. Crit Care Med 2009;37(3):825-832 [FREE Full text] [doi: 10.1097/CCM.0b013e31819b8608] [Medline: 19237884]
20. Ginestra JC, Giannini HM, Schweickert WD, Meadows L, Lynch MJ, Pavan K, et al. Clinician perception of a machine learning-based early warning system designed to predict severe sepsis and septic shock. Crit Care Med 2019;47(11):1477-1484 [FREE Full text] [doi: 10.1097/CCM.0000000000003803] [Medline: 31135500]
21. Daniulaityte R, Zatreh MY, Lamy FR, Nahhas RW, Martins SS, Sheth A, et al. A Twitter-based survey on marijuana concentrate use. Drug Alcohol Depend 2018;187:155-159 [FREE Full text] [doi: 10.1016/j.drugalcdep.2018.02.033] [Medline: 29669296]
22. Eysenbach G. Improving the quality of web surveys: the checklist for reporting results of internet e-surveys (CHERRIES). J Med Internet Res 2004;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
23. Wardi G, Joel I, Villar J, Lava M, Gross E, Tolia V, et al. Equipoise in appropriate initial volume resuscitation for patients in septic shock with heart failure: results of a multicenter clinician survey. J Intensive Care Med 2020;35(11):1338-1345 [FREE Full text] [doi: 10.1177/0885066619871247] [Medline: 31446829]
24. Longoni C, Bonezzi A, Morewedge CK. Resistance to medical artificial intelligence. J Consum Res 2019;46(4):629-650. [doi: 10.1093/jcr/ucz013]

Abbreviations

AI: artificial intelligence
CHERRIES: Checklist for Reporting Results of Internet E-Surveys
ICU: intensive care unit
ML: machine learning
UCSD: University of California, San Diego

XSL•FO
RenderX

Edited by T Leung; submitted 14.07.22; peer-reviewed by C Or, B Barry, X Zhu; comments to author 28.10.22; revised version received 17.11.22; accepted 23.12.22; published 27.01.23.

Please cite as:
Mlodzinski E, Wardi G, Viglione C, Nemati S, Crotty Alexander L, Malhotra A
Assessing Barriers to Implementation of Machine Learning and Artificial Intelligence–Based Tools in Critical Care: Web-Based Survey Study
JMIR Perioper Med 2023;6:e41056
URL: https://periop.jmir.org/2023/1/e41056
doi:10.2196/41056
PMID:36705960

©Eric Mlodzinski, Gabriel Wardi, Clare Viglione, Shamim Nemati, Laura Crotty Alexander, Atul Malhotra. Originally published in JMIR Perioperative Medicine (http://periop.jmir.org), 27.01.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Perioperative Medicine, is properly cited. The complete bibliographic information, a link to the original publication on http://periop.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Temporal Generalizability of Machine Learning Models for Predicting Postoperative Delirium Using Electronic Health Record Data: Model Development and Validation Study

Koutarou Matsumoto[1], MPH, PhD; Yasunobu Nohara[2], PhD; Mikako Sakaguchi[3], BSN; Yohei Takayama[3], MS; Syota Fukushige[4], MS; Hidehisa Soejima[5], MD, PhD; Naoki Nakashima[6], MD, PhD; Masahiro Kamouchi[7,8], MD, PhD

[1]Biostatistics Center, Kurume University, Kurume, Japan
[2]Big Data Science and Technology, Faculty of Advanced Science and Technology, Kumamoto University, Kumamoto, Japan
[3]Department of Nursing, Saiseikai Kumamoto Hospital, Kumamoto, Japan
[4]Department of Inspection, Saiseikai Kumamoto Hospital, Kumamoto, Japan
[5]Institute for Medical Information Research and Analysis, Saiseikai Kumamoto Hospital, Kumamoto, Japan
[6]Medical Information Center, Kyushu University Hospital, Fukuoka, Japan
[7]Department of Health Care Administration and Management, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan
[8]Center for Cohort Studies, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan

Corresponding Author:
Koutarou Matsumoto, MPH, PhD
Biostatistics Center
Kurume University
67 Asahi-Machi
Kurume, 830-0011
Japan
Phone: 81 8033589122
Email: matsumoto_koutarou@kurume-u.ac.jp

Abstract

Background: Although machine learning models demonstrate significant potential in predicting postoperative delirium, the advantages of their implementation in real-world settings remain unclear and require a comparison with conventional models in practical applications.

Objective: The objective of this study was to validate the temporal generalizability of decision tree ensemble and sparse linear regression models for predicting delirium after surgery compared with that of the traditional logistic regression model.

Methods: The health record data of patients hospitalized at an advanced emergency and critical care medical center in Kumamoto, Japan, were collected electronically. We developed a decision tree ensemble model using extreme gradient boosting (XGBoost) and a sparse linear regression model using least absolute shrinkage and selection operator (LASSO) regression. To evaluate the predictive performance of the model, we used the area under the receiver operating characteristic curve (AUROC) and the Matthews correlation coefficient (MCC) to measure discrimination and the slope and intercept of the regression between predicted and observed probabilities to measure calibration. The Brier score was evaluated as an overall performance metric. We included 11,863 consecutive patients who underwent surgery with general anesthesia between December 2017 and February 2022. The patients were divided into a derivation cohort before the COVID-19 pandemic and a validation cohort during the COVID-19 pandemic. Postoperative delirium was diagnosed according to the confusion assessment method.

Results: A total of 6497 patients (68.5, SD 14.4 years, women n=2627, 40.4%) were included in the derivation cohort, and 5366 patients (67.8, SD 14.6 years, women n=2105, 39.2%) were included in the validation cohort. Regarding discrimination, the XGBoost model (AUROC 0.87-0.90 and MCC 0.34-0.44) did not significantly outperform the LASSO model (AUROC 0.86-0.89 and MCC 0.34-0.41). The logistic regression model (AUROC 0.84-0.88, MCC 0.33-0.40, slope 1.01-1.19, intercept –0.16 to 0.06, and Brier score 0.06-0.07), with 8 predictors (age, intensive care unit, neurosurgery, emergency admission, anesthesia time, BMI, blood loss during surgery, and use of an ambulance) achieved good predictive performance.

XSL•FO
RenderX

Conclusions: The XGBoost model did not significantly outperform the LASSO model in predicting postoperative delirium. Furthermore, a parsimonious logistic model with a few important predictors achieved comparable performance to machine learning models in predicting postoperative delirium.

(JMIR Perioper Med 2023;6:e50895) doi:10.2196/50895

KEYWORDS

postoperative delirium; prediction model; machine learning; temporal generalizability; electronic health record data

Introduction

Delirium occurs in a significant proportion of surgical patients, ranging from 11% to 51% [1]. The risk of postoperative delirium is particularly high among patients who receive general anesthesia [1]. Furthermore, postoperative delirium can lead to a prolonged hospital stay and eventually increase mortality [2,3]. Therefore, it is crucial to accurately estimate the potential risk of postoperative delirium in order to identify high-risk patients prior to surgery.

Various tools have been developed thus far to predict postoperative delirium [4-8]. However, the performance of these risk prediction models is not necessarily sufficient, and it is often difficult to collect the data required for some of the predictors used in these models during routine clinical practice. Moreover, study populations differed depending on the type and timing (planned or emergent) of the surgery. In fact, a previous systematic review and meta-analysis concluded that existing models provide weak evidence and thus are not recommended for clinical practice [5]. Recently, machine learning models have garnered attention in the medical field for their high performance in predicting adverse events [9-32]. Thus, machine learning techniques have been recognized as promising tools for the prediction of postoperative delirium [9-17].

Although data-driven models hold great promise for the future, their implementation in real-world settings remains challenging for several reasons. First, it takes time and effort to monitor outliers and missing values in a large amount of data. Therefore, the implementation of complex machine learning models that rely on hundreds of predictors would be impractical in routine clinical practice. Second, it remains unclear whether complex machine learning models offer any additional value compared to conventional tools in practical applications. Although machine learning models can detect complex interactions and nonlinear relationships between predictors and clinical outcomes, some studies have indicated that these complex machine learning models have limited external validity when compared to traditional logistic regression models [33-35]. The “no free lunch” theorem demonstrates that achieving unbiased models that are highly accurate for all data may not be feasible [36]. Thus, it is still unclear whether a complicated model without interpretability is truly superior to a specific model for a particular problem.

The lack of external validity is a significant barrier to implementing predictive models in clinical practice. Additionally, the accuracy of risk prediction models can deteriorate over time due to covariate shifts. For example, during the COVID-19 pandemic, intensive care unit admission was restricted to isolated patients with COVID-19 and care processes were modified according to the measures to reduce the risk of COVID-19 infection. Consequently, models that were developed to predict postoperative delirium before the COVID-19 pandemic are likely to have had reduced predictive accuracy during the pandemic. Assuming that these factors affect predictive performance, complex machine learning models may also not be useful to predict postoperative delirium in patients hospitalized during the pandemic. Therefore, it would be worth evaluating model performance by temporal validation before and after the COVID-19 pandemic.

Thus, this study aimed to determine (1) whether a complex decision tree ensemble model outperforms a sparse linear regression model in predicting postoperative delirium, and (2) whether machine learning prediction models using copious data are superior to traditional regression models with prespecified predictors. For this purpose, we electronically collected existing data from patients who underwent surgery with general anesthesia at a single hospital in Kumamoto, Japan. We developed machine learning models and a traditional linear regression model using data from a cohort of patients hospitalized before the COVID-19 pandemic and compared their performance in predicting postoperative delirium using data in a cohort of patients hospitalized during the COVID-19 pandemic.

Methods

Study Design and Data Sources

We retrospectively reviewed clinical data obtained from patients hospitalized at Saiseikai Kumamoto Hospital, which is designated as an advanced emergency and critical care medical center in Kumamoto City in the southern region of Japan. We accessed, processed, and analyzed clinical data that had been electronically stored within the hospital’s database. We developed the prediction models in accordance with the guidelines outlined in the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis statement [37] (Multimedia Appendix 1).

Ethics Approval

This study was approved by the institutional review board of Saiseikai Kumamoto Hospital (approval number 1072). Owing to the retrospective nature of the study and the use of anonymized clinical data, the requirement for written informed consent was waived. For patients who did not wish to participate in this study, the opportunity to opt out was provided and announced on the Saiseikai Kumamoto Hospital website. During the analysis, patient data were deidentified, and only

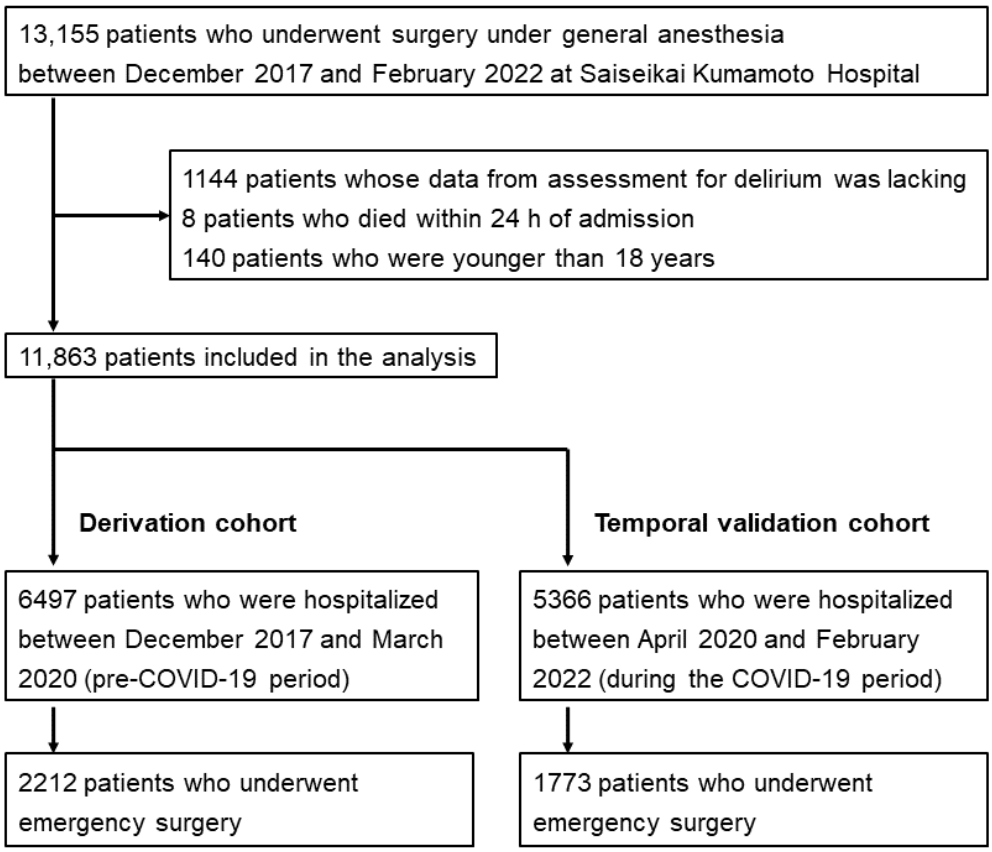
XSL•FO
RenderX

anonymized information was used. Participation in this study was voluntary, and participants were not compensated for the use of their information.

Study Patients

We obtained data for a total of 13,155 patients who were admitted to the hospital between December 2017 and February 2022, underwent surgery with general anesthesia, and were discharged before February 2022. Among these patients, we excluded 1144 patients due to missing data on delirium assessment, 8 patients who died within 24 hours of admission, and 140 patients who were younger than 18 years. Finally, a total of 11,863 patients were included in the analysis. A subanalysis was conducted by restricting the patient population to those who underwent emergent surgery (Figure 1).

Figure 1. Patient selection flowchart.



Clinical Outcomes

Delirium was diagnosed according to the confusion assessment method (CAM) [38], which is the most widely used instrument for diagnosing delirium [1]. We defined postoperative delirium as the first positive result obtained during hospitalization after surgery using the CAM. Each patient underwent daily assessments using the CAM, starting from the day they were transferred from the operating room to the postoperative care unit. The assessments were conducted by nurses who received training from a delirium assessment committee, which included a psychiatrist.

Predictors

We collected data on predictors from electronic health record databases, which were routinely accumulated during daily patient care at the hospital. To develop prediction models for patients undergoing surgery under general anesthesia, we used 25 variables that were collected on admission (eg, age, BMI, sex, emergency admission, use of an ambulance, medication, comorbidity, and previous history) as well as during the preoperative (eg, admission ward, catheter, ventilator, physical restraints, and circadian rhythm disorder) and intraoperative (eg, surgery site, anesthesia time, and blood loss during surgery) periods. All variables used for prediction are listed in Multimedia Appendix 2. Additionally, to develop the prediction models for patients who underwent emergent surgery under general anesthesia, we used 30 additional variables, including data derived from blood tests (eg, albumin, creatinine, glucose, hemoglobin, and C-reactive protein levels) and vital sign assessments (eg, systolic blood pressure, pulse rate, SpO_2, and Glasgow Coma Scale score) upon admission (Multimedia Appendix 2).

Development of Predictive Models

The extreme gradient boosting (XGBoost) effectively captures nonlinear relationships and interactions between predictors and outcomes [39]. Based on these characteristics, previous studies have reported that XGBoost exhibits a high level of accuracy in predicting postoperative delirium [11-15]. Therefore, we used XGBoost, which is a decision tree ensemble learning method, as a complex machine learning model. Additionally, we also

XSL•FO
RenderX

used the least absolute shrinkage and selection operator (LASSO) regression, a sparse linear regression method [40]. LASSO regression serves as a variable selection technique by eliminating regression coefficients of variables that do not significantly contribute to the prediction. To create a machine learning–referenced regression model, we further used logistic regression with prespecified predictors that were identified as the top 10 important variables in both the XGBoost and LASSO models.

The data set contained missing values for several predictors (Multimedia Appendix 2). While multiple imputations are considered the gold standard for handling missing data, its compatibility with decision-tree–based machine learning models is not as straightforward as it is with regression models. Therefore, we used the MissForest method, a type of nonparametric single-imputation approach, for missing value imputation [41]. This method uses other variables, including the outcome, to predict the missing values of the predictor variable using random forest. Since the rate of missing values was at most 8.5%, we used all variables with missing values after imputing them using MissForest. The derivation and the validation cohorts were handled as completely independent data sets, and missing values were imputed within each data set. Additionally, for sensitivity analysis, we conducted a complete case analysis to assess the impact of missing values on the results. Further details regarding the programs used, along with the GitHub URL, are outlined in Multimedia Appendix 3.

Generalizability Assessment

To verify the general applicability of the prediction model, we performed temporal validation, which is a form of external validation [42,43]. To explore the temporal generalizability, we divided the patients into 2 cohorts based on whether they were admitted to the hospital before or after the onset of the COVID-19 pandemic in Japan. The derivation cohort consisted of patients hospitalized between December 2017 and March 2020, while the validation cohort comprised patients hospitalized between April 2020 and February 2022 (Figure 1).

We developed prediction models using the derivation cohort and then validated the models using the temporal validation cohort. When developing the predictive models in the derivation cohort, the internal validity of the predictive models was assessed using stratified 5-fold cross-validation. The discrimination performance was evaluated based on sensitivity, specificity, positive predictive value, negative predictive value, and area under the receiver operating characteristic curve (AUROC), while the calibration performance was assessed using the calibration slope and intercept of the regression line between the predicted and observed probabilities. The overall performance of the models was measured using the Brier score.

Furthermore, the Matthews correlation coefficient (MCC) and the area under the precision-recall curve (AUPRC) were calculated to provide a fair assessment of binary classification for biased data, such as cases with a small number of events. The MCC takes values between –1 and +1, with a value of +1 indicating perfect agreement between the predicted and observed values, –1 indicating perfect disagreement between the predicted and observed values, and 0 indicating equivalent performance to random prediction [44]. The MCC was defined as follows:

In the equation, TP stands for true positives, TN for true negatives, FP for false positives, and FN for false negatives. For each metric, the 95% CI was calculated using 2000 bootstrap samples.

Contribution of Predictors to Predictive Performance

We calculated the Shapley additive explanations (SHAP) values to assess the contribution of each predictor in the XGBoost model. SHAP is a model-agnostic machine learning interpretability method that provides a valuable tool for visualizing the contribution of predictors using Shapley values. These values, derived from cooperative game theory, ensure a fair distribution of the contribution across predictors [45]. In the LASSO regression model, we estimated the standardized partial regression coefficients to assess the contribution of each predictor. To evaluate the improvement in predictive performance attributed to each predictor, we compared the AUROC, calibration slope, and calibration intercept after sequentially adding the predictors to the models. The predictive metrics were calculated as each predictor was implemented in the model in the order of its predictive contribution.

Statistical Analysis

To compare the baseline characteristics, we used the χ^2 test or Fisher exact test for categorical data, the *t* test for data assumed to follow a normal distribution, and the Mann-Whitney *U* test for continuous values when a normal distribution could not be assumed. The Kolmogorov-Smirnov test was used to determine whether the continuous variable data followed the normal distribution. The *t* test was used to evaluate differences in age and BMI between the 2 groups. The 2-sided probability values less than .05 were considered statistically significant. All statistical analyses were conducted using the R statistical package (R Core Team) (Multimedia Appendix 3).

Results

Patient Characteristics

The derivation cohort consisted of 6497 patients with a mean age of 68.5 (SD 14.4) years of which 2627 (40.4%) were women. The validation cohort included 5366 patients with a mean age of 67.8 (SD 14.6) years of which 2105 (39.2%) were women. Postoperative delirium occurred in 592 (9.1%) patients in the derivation cohort and in 427 (8%) patients in the validation cohort. Delirium developed within 3 days after surgery in 66% of patients, within a week after surgery in 84% of patients, and within 2 weeks after surgery in 94% of patients.

Table 1 shows the differences in the baseline characteristics between patients with and without postoperative delirium in both the derivation and validation cohorts. In the derivation cohort, all predictors significantly differed, except for central venous port, dialysis catheter, and circadian rhythm disorder. Similar results were observed in the validation cohort, although the frequency of the dialysis catheter differed.

XSL•FO
RenderX

Table 1. Differences in the baseline data according to the absence or presence of delirium after surgery.

	Derivation cohort		*P* value	Validation cohort		*P* value
	No delirium (n=5905)	Delirium (n=592)		No delirium (n=4939)	Delirium (n=427)	
Patient data						
Age (years), mean (SD)	67.8 (14.3)	76.0 (12.7)	<.001	67.0 (14.5)	78.0 (12.5)	<.001
BMI (kg/m²), mean (SD)	23.2 (3.9)	21.9 (3.8)	<.001	23.3 (4.0)	21.3 (3.8)	<.001
Women, n (%)	2358 (39.9)	269 (45.4)	.01	1912 (38.7)	193 (45.2)	.01
Emergency admission, n (%)	1814 (30.7)	398 (67.2)	<.001	1474 (29.8)	299 (70.0)	<.001
Use of ambulance, n (%)	1110 (18.8)	326 (55.1)	<.001	877 (17.8)	252 (59.0)	<.001
Medication, n (%)						
Benzodiazepines	93 (1.7)	18 (3.2)	.02	63 (1.3)	14 (3.3)	.002
Opioids	55 (1.0)	29 (5.1)	<.001	40 (0.8)	28 (6.6)	<.001
Steroids	62 (1.2)	20 (3.5)	<.001	54 (1.1)	17 (4.0)	<.001
Dementia, n (%)	257 (4.8)	88 (15.5)	<.001	176 (3.6)	115 (26.9)	<.001
Brain disease, n (%)	509 (9.5)	118 (20.8)	<.001	383 (7.8)	80 (18.7)	<.001
Previous history, n (%)						
Heavy drinking	47 (0.9)	12 (2.1)	.009	45 (0.9)	12 (2.8)	.001
Delirium	84 (1.6)	32 (5.7)	<.001	94 (1.9)	55 (12.9)	<.001
Preoperative data						
Admission ward, n (%)			<.001			<.001
General ward (shared room)	2085 (35.3)	123 (20.8)		1666 (33.7)	77 (18.0)	
General ward (private room)	2940 (49.8)	156 (26.4)		2660 (53.9)	131 (30.7)	
Intensive care unit	880 (14.9)	313 (52.9)		613 (12.4)	219 (51.3)	
Catheter, n (%)						
Indwelling urinary catheter	1663 (28.2)	288 (48.6)	<.001	1110 (22.5)	209 (48.9)	<.001
Peripheral vein catheter	2856 (48.4)	423 (71.5)	<.001	2136 (43.2)	317 (74.2)	<.001
Central venous catheter	322 (5.5)	96 (16.2)	<.001	241 (4.9)	77 (18.0)	<.001
Central venous port	39 (0.7)	4 (0.7)	>.99	25 (0.5)	3 (0.7)	.49
Dialysis catheter	45 (0.8)	9 (1.5)	.09	36 (0.7)	13 (3.0)	<.001
Swan-Ganz catheter	143 (2.4)	47 (7.9)	<.001	81 (1.6)	45 (10.5)	<.001
Ventilator, n (%)	4341 (73.5)	500 (84.5)	<.001	3259 (66.0)	351 (82.2)	<.001
Physical restraints, n (%)	124 (2.1)	41 (6.9)	<.001	71 (1.4)	40 (9.4)	<.001
Circadian rhythm disorder, n (%)	19 (0.3)	2 (0.3)	>.99	27 (0.5)	4 (0.9)	.31
Surgical data						
Surgery site, n (%)			<.001			<.001
Thoracic cavity and mediastinum	499 (8.5)	7 (1.2)		430 (8.7)	4 (0.9)	
Chest wall, abdominal wall, and perineum	294 (5.0)	6 (1.0)		264 (5.3)	10 (2.3)	
Upper abdominal viscera	1421 (24.1)	112 (18.9)		1047 (21.2)	70 (16.4)	
Lower abdominal viscera	1526 (25.8)	117 (19.8)		1315 (26.6)	107 (25.1)	
Hip joints and extremities	917 (15.5)	83 (14.0)		1017 (20.6)	78 (18.3)	
Central nervous system	584 (9.9)	114 (19.3)		358 (7.2)	43 (10.1)	
Heart and vascular	603 (10.2)	150 (25.3)		480 (9.7)	114 (26.7)	
Other	61 (1.0)	3 (0.5)		28 (0.6)	1 (0.2)	

XSL•FO
RenderX

	Derivation cohort		*P* value	Validation cohort		*P* value
	No delirium (n=5905)	Delirium (n=592)		No delirium (n=4939)	Delirium (n=427)	
Anesthesia time, min, median (IQR)	209 (137-309)	236 (156-403)	<.001	197 (136-305)	206 (146-379)	<.001
Blood loss during surgery (mL), median (IQR)	20 (5-150)	120 (5-703)	<.001	20 (5-100)	70 (10-600)	<.001

Predictive Performance of Machine Learning Models

The performance of the prediction models developed using the derivation cohort was first evaluated using stratified 5-fold cross-validation. As a result, the AUROC values for the XGBoost, LASSO, and logistic regression models were found to be 0.85 (SD 0.02), 0.85 (SD 0.02), and 0.854 (SD 0.02), respectively. Thereafter, we validated the performance using the temporal validation cohort of patients hospitalized during the COVID-19 pandemic. The predictive performance of the machine learning models is shown in Table 2. In terms of discrimination performance, the LASSO model showed comparable results to the XGBoost model, with AUROC values ranging 0.86-0.89 and 0.87-0.90, sensitivity of 0.80-0.90 and 0.77-0.91, specificity of 0.73-0.82 and 0.73-0.85, positive predictive value of 0.22-0.28 and 0.22-0.31, and negative predictive value of 0.98-0.99 and 0.98-0.99, respectively. Similarly, MCC values and the AUPRC for the LASSO model ranged from 0.34 to 0.41 and from 0.36 to 0.46, respectively, while the XGBoost model showed values from 0.34 to 0.44 and from 0.36 to 0.46, respectively. Regarding calibration, the LASSO model exhibited good calibration with slope values ranging from 0.98 to 1.14. On the other hand, the XGBoost model displayed a more prominent slope greater than 1, ranging from 1.13 to 1.32, indicating suboptimal calibration of the decision tree ensemble model. Additionally, both the XGBoost and LASSO models demonstrated comparable performance in terms of Brier score, with values ranging from 0.05 to 0.06 for both models.

Table 2. Temporal validation of predictive models for delirium after surgery[a].

	XGBoost[b]	LASSO[c]	LR[d]
Discriminability metrics			
AUROC[e] (95% CI)	0.88 (0.87 to 0.90)	0.88 (0.86 to 0.89)	0.86 (0.84 to 0.88)
Sensitivity (95% CI)	0.88 (0.77 to 0.91)	0.84 (0.80 to 0.90)	0.84 (0.75 to 0.87)
Specificity (95% CI)	0.74 (0.73 to 0.85)	0.78 (0.73 to 0.82)	0.75 (0.74 to 0.84)
PPV[f] (95% CI)	0.23 (0.22 to 0.31)	0.25 (0.22 to 0.28)	0.23 (0.22 to 0.29)
NPV[g] (95% CI)	0.99 (0.98 to 0.99)	0.98 (0.98 to 0.99)	0.98 (0.97 to 0.99)
Discriminability metrics for an imbalanced event			
MCC[h] (95% CI)	0.37 (0.34 to 0.44)	0.38 (0.34 to 0.41)	0.35 (0.33 to 0.40)
AUPRC[i] (95% CI)	0.41 (0.36 to 0.46)	0.41 (0.36 to 0.46)	0.35 (0.31 to 0.40)
Calibration metrics			
Slope (95% CI)	1.22 (1.13 to 1.32)	1.06 (0.98 to 1.14)	1.10 (1.01 to 1.19)
Intercept (95% CI)	–0.05 (–0.16 to 0.06)	–0.26 (–0.38 to –0.15)	–0.05 (–0.16 to 0.06)
Overall metric			
Brier score (95% CI)	0.06 (0.05 to 0.06)	0.06 (0.05 to 0.06)	0.06 (0.06 to 0.07)

[a]The predictive models were developed on the training cohorts and validated on the test cohorts. A logistic regression model was developed using the key predictors identified by the machine learning models: age, intensive care unit, neurosurgery, emergency admission, anesthesia time, BMI, blood loss during surgery, and use of an ambulance. The values in parentheses represent 95% CIs after 2000 bootstrap samples.

[b]XGBoost: extreme gradient boosting.

[c]LASSO: least absolute shrinkage and selection operator.

[d]LR: logistic regression.

[e]AUROC: area under the receiver operating characteristic curve.

[f]PPV: positive predictive value.

[g]NPV: negative predictive value.

[h]MCC: Matthews correlation coefficient.

[i]AUPRC: area under the precision-recall curve.

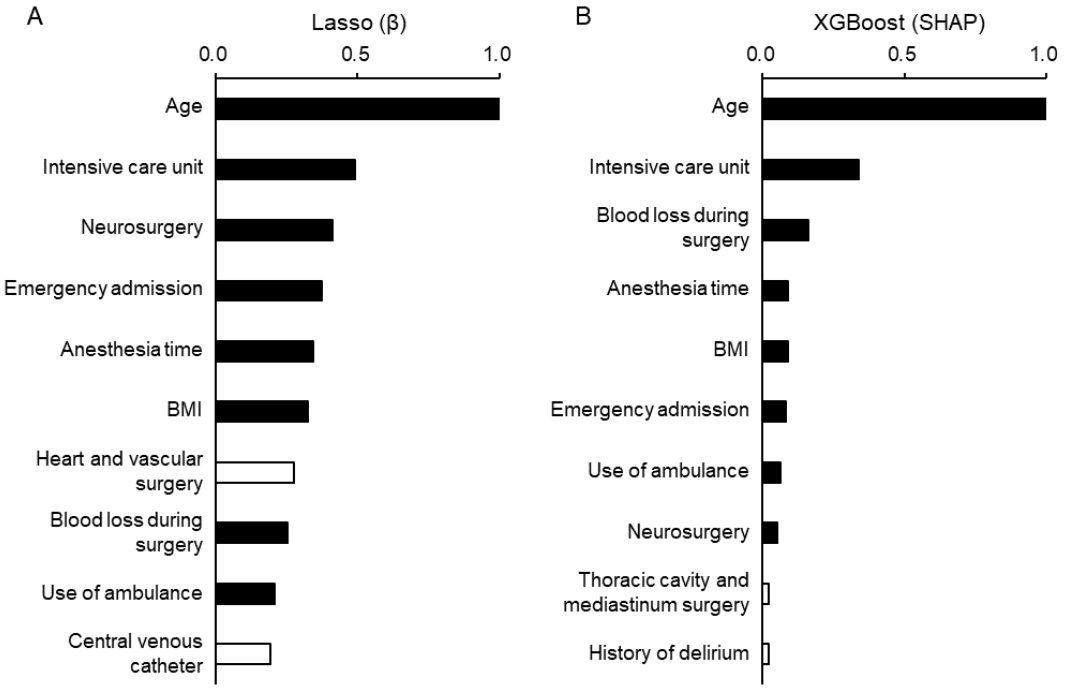
XSL•FO
RenderX

We also validated the predictive performance of the logistic regression model using 8 key predictors—age, intensive care unit, neurosurgery, emergency admission, anesthesia time, BMI, blood loss during surgery, and use of an ambulance—that were identified as important in both the LASSO and XGBoost models (Table 2). The results showed that the logistic regression model also exhibited excellent performance compared with the machine learning models. Discriminability metrics (AUROC 0.84-0.88, sensitivity 0.75-0.87, specificity 0.74-0.84, positive predictive value 0.22-0.29, and negative predictive value 0.97-0.99), discriminability metrics for an imbalanced event (MCC 0.33-0.40 and AUPRC 0.31-0.40), calibration metrics (slope 1.01-1.19 and intercept –0.16 to 0.06), and overall metric (Brier score 0.06-0.07) were comparable to those of the machine learning models.

Factors Contributing to Prediction

Figure 2 illustrates the variable importance of the top 10 predictors identified in both the LASSO and XGBoost models. The ranking of variable importance was similar for each variable in both models. However, some of the high-ranked variables in the LASSO model, such as heart and vascular surgery and central venous catheter, were not necessarily regarded as important in the XGBoost model.

Figure 2. Variable importance for delirium after surgery in the machine learning models. The graph shows the variable importance of the top 10 ranked predictors in the (A) LASSO and (B) XGBoost models. Variable importance was assessed based on the standardized regression coefficient (β) for the LASSO model and the SHAP value for the XGBoost model and is depicted as the value relative to the highest value. Solid bars indicate the variables that are ranked in the top 10 for both models. LASSO: least absolute shrinkage and selection operator; SHAP: Shapley additive explanations; XGBoost: extreme gradient boosting.



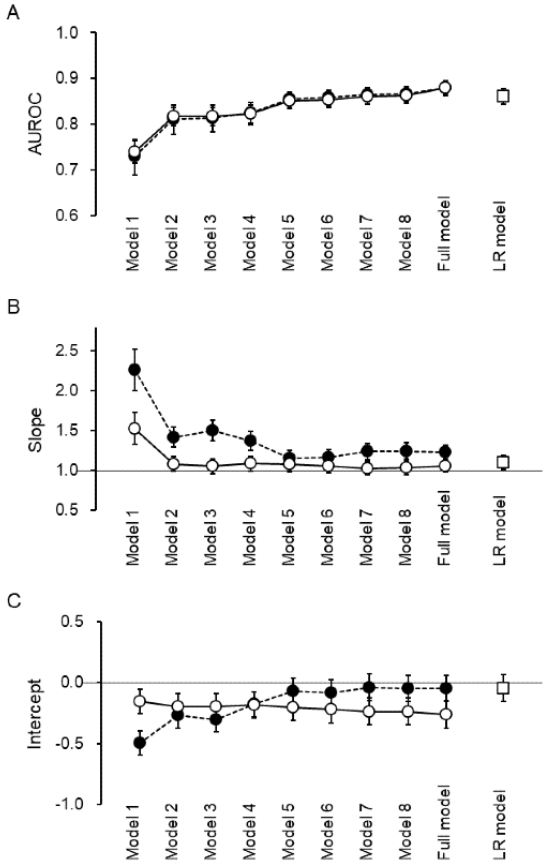
Predictors and Predictive Performance

Figure 3 demonstrates the impact of including additional predictors on discrimination and calibration. The AUROC did not improve when low-ranked predictors were added to a set of high-ranked predictors in both the LASSO and XGBoost models (Figure 3A).

When calibration is perfect, the slope and the intercept of the calibration plot should be 1 and 0, respectively. The slope in the LASSO model approached 1, whereas the slope in the XGBoost model did not reach 1 even with an increase in the number of predictors (Figure 3B). On the other hand, the intercept in the XGBoost model approached 0 as the number of predictors increased, while it decreased below 0 in the LASSO model (Figure 3C).

XSL•FO
RenderX

Figure 3. Predictive performance for delirium after surgery of machine learning models and the traditional logistic regression model. The (A) AUROC, (B) calibration slope, and (C) calibration intercept with the increase in the number of predictors are shown for the XGBoost model (closed circles) and LASSO model (open circles) in comparison with those for the logistic regression model using prespecified predictors (open squares). The predictors used in the logistic regression model were those ranked in the top 10 for both the XGBoost and LASSO models. The predictors used in each model are as follows: Model 1: age; Model 2: model 1 + intensive care unit; Model 3: model 2 + neurosurgery; Model 4: model 3 + emergency admission; Model 5: model 4 + anesthesia time; Model 6: model 5 + BMI; Model 7: model 6 + blood loss during surgery; Model 8: model 7 + use of an ambulance; Full model: all 25 variables. AUROC: area under the receiver operating characteristic curve; LASSO: least absolute shrinkage and selection operator; LR: logistic regression; XGBoost: extreme gradient boosting.



Comparison Between Machine Learning Models and the Logistic Regression Model

We further compared the predictive performance between the machine learning models and the traditional logistic regression model using prespecified predictors. For the logistic regression model, we selected 8 predictors: age, intensive care unit, neurosurgery, emergency admission, anesthesia time, BMI, blood loss during surgery, and use of an ambulance. These predictors were among the top 10 variables associated with postoperative delirium identified by the XGBoost and LASSO models. The logistic regression model, incorporating these 8 predictors, demonstrated sufficient discriminative ability (AUROC 0.84-0.88 and MCC 0.33-0.40), which was comparable to the machine learning models that used all available data (Table 2, Figure 3A). Furthermore, logistic regression with predictors selected based on the machine learning models exhibited better calibration (slope 1.01-1.19 and intercept –0.16 to 0.06) than the XGBoost model (Table 2, Figure 3B). In terms of overall performance (Brier score 0.06-0.07), the logistic regression model was comparable to the machine learning models (Table 2).

Sensitivity Analysis

We conducted a sensitivity analysis by restricting the study patients to those who underwent emergent surgery (Multimedia Appendix 4). As a result, similar trends were confirmed in these patients. The discriminative ability between the LASSO and XGBoost models was comparable, whereas the XGBoost model showed poor calibration (Multimedia Appendices 5 and 6). Among the top 10 predictors identified by both the XGBoost and LASSO models, 5 predictors (age, intensive care unit, Glasgow Coma Scale score, anesthesia time, and blood loss during surgery) were found in both models (Multimedia Appendix 6). Consequently, the logistic regression model, incorporating these 5 predictors, demonstrated comparable discriminative ability to the machine learning models, while exhibiting better calibration than the XGBoost model (Multimedia Appendices 5 and 6).

Finally, we performed a complete case analysis and found that the results were consistent with the main analysis. The discriminative power among the XGBoost, LASSO, and the logistic regression model was comparable, while the calibration of the XGBoost model was inferior to that of the logistic regression model (Multimedia Appendix 5).

XSL•FO
RenderX

Discussion

Principal Results

The temporal generalizability analysis revealed that machine learning exhibited a high discriminative ability in predicting postoperative delirium in a real-world setting. However, increasing the number of predictors did not considerably improve the discriminative performance, even for the machine learning models. The complex ensemble decision tree model did not outperform the sparse linear regression model in terms of discriminative power, and it exhibited poor calibration. In contrast, the traditional logistic regression model with a limited number of important predictors achieved sufficient discriminatory ability in predicting postoperative delirium and demonstrated better calibration than the complex ensemble decision tree model. These findings suggest that a traditional model with prespecified important predictors would be more practical and useful in estimating the risk of postoperative delirium compared to machine learning models.

Comparison With Prior Work

In this study, we validated the prediction models in terms of temporal generalizability before and after the COVID-19 pandemic, as the health care system underwent significant changes following the pandemic. The pandemic led to a decrease in the number of hospital admissions and treatments for various diseases [46,47].

Additionally, the use of hospital wards, particularly intensive care units, was modified for COVID-19 prevention and control measures. Therefore, we had concerns that the models, especially complex models such as decision tree ensemble models, would overfit the derivation cohort and thus not be applicable to the validation cohort if many predictors were included without a priori verification of the changes in the predictors. Nevertheless, both the decision tree ensemble model and sparse linear regression model exhibited high discriminative power for predicting postoperative delirium. Therefore, machine learning appears to be a useful technique only if we simply categorize patients into high-risk and low-risk groups based on the risk of delirium after surgery using comprehensive data.

In contrast, several studies have shown that machine learning models do not necessarily outperform linear regression models in terms of calibration and generalizability, suggesting that linear regression models may be sufficient in low-dimensional settings with large data sets [33-35]. Furthermore, recent simulation studies on material science data sets have suggested that simple linear regression models are preferable to complex machine learning models, such as a random forest, in terms of extrapolated predictive performance [48]. Our results are consistent with this notion as the linear regression model using only the key predictors accurately predicted postoperative delirium in patients who underwent surgery under general anesthesia. In real-world settings, it would be challenging to manage a large number of predictors, incorporate them into predictive models, and implement prediction systems for postoperative delirium. Instead, it would be convenient if a smaller number of key predictors could provide sufficient accuracy in predicting delirium after surgery. Thus, the implementation of a parsimonious linear regression model rather than a complex machine learning model may be practical and useful in predicting the risk of postoperative delirium.

Although the discriminative ability of the machine learning models and the logistic regression model was comparable, the traditional logistic regression model exhibited better calibration performance than the XGBoost model. In daily clinical practice, poorly calibrated risk estimates can lead to incorrect strategies for safeguarding against postoperative delirium. In this respect, the calibration of the predictive model is crucial for estimating risk for each patient. Flexible models, such as boosted trees, have been reported to be poorly calibrated, and their output prediction probabilities are often corrected using sigmoid functions or isotonic regression [49]. Nevertheless, such a complex conversion process is cumbersome and not cost-effective if prediction systems are implemented in clinical practice.

The implementation of machine learning in real-world settings has gained increasing attention in recent research. Machine learning models have the potential to outperform linear regression models in predicting clinical outcomes when there are nonlinear relationships between the major predictors and the outcomes or when strong interactions exist among the predictors. In such complex scenarios, the predictive accuracy of a flexible model, such as a decision tree ensemble, may surpass that of a linear model. Furthermore, the decision tree ensemble model can effectively handle high-dimensional data without the issues of multicollinearity inherent in linear regression models. Consequently, machine learning techniques offer valuable tools for exploring and identifying important predictors among numerous variables. The selection of appropriate predictive models requires careful consideration of their advantages and disadvantages. Further theoretical and empirical studies are needed to ascertain the use of machine learning models in predicting clinical outcomes in daily clinical practice.

Limitations

This study had several limitations to be considered. First, some variables contained missing values. Although imputation techniques were used, the results may have been affected—specifically, missing predictor values were imputed using all available data in MissForest. Nevertheless, even when the outcome was excluded in predicting missing values using the imputation method, the results were essentially the same (data not shown). Second, relying on a limited number of predictors obtained during routine clinical work leaves the possibility of unmeasured predictors influencing the outcomes. Third, the analysis was conducted using data from a single center in Japan, necessitating validation of the findings in other settings. Finally, variations in health policies during the COVID-19 pandemic across countries warranted confirmation of the temporal validity in different contexts.

Conclusions

In recent years, the application of data-driven models to electronic medical record data for outcome prediction has garnered interest. However, the focus has primarily been on

XSL•FO
RenderX

complex state-of-the-art machine learning models. In our study, the decision tree ensemble model demonstrated a comparable discriminatory ability to a sparse linear regression model in predicting postoperative delirium using real-world data. Additionally, a simple traditional logistic regression model using known predictors outperformed the complex ensemble decision tree model in terms of calibration. Therefore, a cautious evaluation of the advantages and disadvantages of data-driven models for postoperative delirium prediction is warranted.

Acknowledgments

We thank the staff of Saiseikai Kumamoto Hospital for their help with data collection. This study was supported by Grants-in-Aid for Scientific Research (KAKENHI; grants JP22K17336 and JP22H03328) from the Japanese Ministry of Education and the Ministry of Health, Labour and Welfare (MHLW) Program (grant JP21AC1002) from the Japanese MHLW. Generative artificial intelligence was not used in the writing of this paper. The sponsors had no role in the design and conduct of the study; the collection, management, analysis, or interpretation of the data; preparation, review, or approval of the paper; or decision to submit the paper for publication.

Data Availability

The data sets generated and analyzed during this study are available from the corresponding author upon reasonable request. However, owing to the confidential nature of the data, only qualified researchers who are trained in human subject confidentiality protocols may submit requests to use the data set to the corresponding author.

Authors' Contributions

KM had full access to all the study data and takes responsibility for study data integrity and the accuracy of the data analysis. The study concept and design were handled by KM, MK, NN, and HS; acquisition, analysis, and interpretation of data were undertaken by KM, YN, MS, YT, and SF; the paper was drafted by KM and MK, with MK carrying out critical revision of the paper for important intellectual content; statistical analysis was conducted by KM and YN; funding was obtained by KM and NN; administrative, technical, and material support was provided by MS, YT, and SF; and the project was supervised by MK, HS, and NN.

Conflicts of Interest

None declared.

Multimedia Appendix 1
TRIPOD (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis) checklist.
[DOCX File , 98 KB - periop_v6i1e50895_app1.docx]

Multimedia Appendix 2
Missing values in the baseline data.
[DOCX File , 41 KB - periop_v6i1e50895_app2.docx]

Multimedia Appendix 3
Supplemental analysis methods.
[DOCX File , 21 KB - periop_v6i1e50895_app3.docx]

Multimedia Appendix 4
Differences in baseline data according to the absence or presence of delirium.
[DOCX File , 36 KB - periop_v6i1e50895_app4.docx]

Multimedia Appendix 5
Temporal validation of predictive models for delirium.
[DOCX File , 32 KB - periop_v6i1e50895_app5.docx]

Multimedia Appendix 6
Predictive performance of machine learning models for delirium after emergent surgery in comparison with a traditional logistic regression model. Variable importance for delirium after emergent surgery in the machine learning models.
[DOCX File , 172 KB - periop_v6i1e50895_app6.docx]

XSL•FO
RenderX

References

1. Inouye SK, Westendorp RGJ, Saczynski JS. Delirium in elderly people. Lancet 2014;383(9920):911-922 [FREE Full text] [doi: 10.1016/S0140-6736(13)60688-1] [Medline: 23992774]
2. Witlox J, Eurelings LSM, de Jonghe JFM, Kalisvaart KJ, Eikelenboom P, van Gool WA. Delirium in elderly patients and the risk of postdischarge mortality, institutionalization, and dementia: a meta-analysis. JAMA 2010 Jul 28;304(4):443-451. [doi: 10.1001/jama.2010.1013] [Medline: 20664045]
3. Edlund A, Lundström M, Karlsson S, Brännström B, Bucht G, Gustafson Y. Delirium in older patients admitted to general internal medicine. J Geriatr Psychiatry Neurol 2006;19(2):83-90. [doi: 10.1177/0891988706286509] [Medline: 16690993]
4. Jansen CJ, Absalom AR, de Bock GH, van Leeuwen BL, Izaks GJ. Performance and agreement of risk stratification instruments for postoperative delirium in persons aged 50 years or older. PLoS One 2014;9(12):e113946 [FREE Full text] [doi: 10.1371/journal.pone.0113946] [Medline: 25464335]
5. van Meenen LCC, van Meenen DMP, de Rooij SE, ter Riet G. Risk prediction models for postoperative delirium: a systematic review and meta-analysis. J Am Geriatr Soc 2014;62(12):2383-2390. [doi: 10.1111/jgs.13138] [Medline: 25516034]
6. Kim MY, Park UJ, Kim HT, Cho WH. DELirium prediction based on hospital information (Delphi) in general surgery patients. Medicine (Baltimore) 2016;95(12):e3072 [FREE Full text] [doi: 10.1097/MD.0000000000003072] [Medline: 27015177]
7. Choi NY, Kim EH, Baek CH, Sohn I, Yeon S, Chung MK. Development of a nomogram for predicting the probability of postoperative delirium in patients undergoing free flap reconstruction for head and neck cancer. Eur J Surg Oncol 2017;43(4):683-688 [FREE Full text] [doi: 10.1016/j.ejso.2016.09.018] [Medline: 27773516]
8. Whitlock EL, Braehler MR, Kaplan JA, Finlayson E, Rogers SE, Douglas V, et al. Derivation, validation, sustained performance, and clinical impact of an electronic medical record-based perioperative delirium risk stratification tool. Anesth Analg 2020;131(6):1901-1910 [FREE Full text] [doi: 10.1213/ANE.0000000000005085] [Medline: 33105280]
9. Mufti HN, Hirsch GM, Abidi SR, Abidi SSR. Exploiting machine learning algorithms and methods for the prediction of agitated delirium after cardiac surgery: models development and validation study. JMIR Med Inform 2019;7(4):e14993 [FREE Full text] [doi: 10.2196/14993] [Medline: 31558433]
10. Oosterhoff JHF, Karhade AV, Oberai T, Franco-Garcia E, Doornberg JN, Schwab JH. Prediction of postoperative delirium in geriatric hip fracture patients: a clinical prediction model using machine learning algorithms. Geriatr Orthop Surg Rehabil 2021;12:21514593211062277 [FREE Full text] [doi: 10.1177/21514593211062277] [Medline: 34925951]
11. Bishara A, Chiu C, Whitlock EL, Douglas VC, Lee S, Butte AJ, et al. Postoperative delirium prediction using machine learning models and preoperative electronic health record data. BMC Anesthesiol 2022;22(1):8 [FREE Full text] [doi: 10.1186/s12871-021-01543-y] [Medline: 34979919]
12. Zhang Y, Wan DH, Chen M, Li YL, Ying H, Yao GL, et al. Automated machine learning-based model for the prediction of delirium in patients after surgery for degenerative spinal disease. CNS Neurosci Ther 2023;29(1):282-295 [FREE Full text] [doi: 10.1111/cns.14002] [Medline: 36258311]
13. Xue B, Li D, Lu C, King CR, Wildes T, Avidan MS, et al. Use of machine learning to develop and evaluate models using preoperative and intraoperative data to identify risks of postoperative complications. JAMA Netw Open 2021;4(3):e212240 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.2240] [Medline: 33783520]
14. Wang Y, Lei L, Ji M, Tong J, Zhou CM, Yang JJ. Predicting postoperative delirium after microvascular decompression surgery with machine learning. J Clin Anesth 2020;66:109896 [FREE Full text] [doi: 10.1016/j.jclinane.2020.109896] [Medline: 32504969]
15. Jung JW, Hwang S, Ko S, Jo C, Park HY, Han HS, et al. A machine-learning model to predict postoperative delirium following knee arthroplasty using electronic health records. BMC Psychiatry 2022;22(1):436 [FREE Full text] [doi: 10.1186/s12888-022-04067-y] [Medline: 35761274]
16. Davoudi A, Ebadi A, Rashidi P, Ozrazgat-Baslanti T, Bihorac A, Bursian AC. Delirium prediction using machine learning models on preoperative electronic health records data. Proc IEEE Int Symp Bioinformatics Bioeng 2017;2017:568-573 [FREE Full text] [doi: 10.1109/BIBE.2017.00014] [Medline: 30393788]
17. Zhao H, You J, Peng Y, Feng Y. Machine learning algorithm using electronic chart-derived data to predict delirium after elderly hip fracture surgeries: a retrospective case-control study. Front Surg 2021;8:634629 [FREE Full text] [doi: 10.3389/fsurg.2021.634629] [Medline: 34327210]
18. Wong A, Young AT, Liang AS, Gonzales R, Douglas VC, Hadley D. Development and validation of an electronic health record-based machine learning model to estimate delirium risk in newly hospitalized patients without known cognitive impairment. JAMA Netw Open 2018;1(4):e181018 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.1018] [Medline: 30646095]
19. Matsumoto K, Nohara Y, Soejima H, Yonehara T, Nakashima N, Kamouchi M. Stroke prognostic scores and data-driven prediction of clinical outcomes after acute ischemic stroke. Stroke 2020;51(5):1477-1483 [FREE Full text] [doi: 10.1161/STROKEAHA.119.027300] [Medline: 32208843]
20. Corey KM, Kashyap S, Lorenzi E, Lagoo-Deenadayalan SA, Heller K, Whalen K, et al. Development and validation of machine learning models to identify high-risk surgical patients using automatically curated electronic health record data

XSL•FO
RenderX

(Pythia): a retrospective, single-site study. PLoS Med 2018;15(11):e1002701 [FREE Full text] [doi: 10.1371/journal.pmed.1002701] [Medline: 30481172]
21. Sun X, Douiri A, Gulliford M. Applying machine learning algorithms to electronic health records to predict pneumonia after respiratory tract infection. J Clin Epidemiol 2022;145:154-163 [FREE Full text] [doi: 10.1016/j.jclinepi.2022.01.009] [Medline: 35045315]
22. Tiwari P, Colborn KL, Smith DE, Xing F, Ghosh D, Rosenberg MA. Assessment of a machine learning model applied to harmonized electronic health record data for the prediction of incident atrial fibrillation. JAMA Netw Open 2020;3(1):e1919396 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.19396] [Medline: 31951272]
23. Rose S. Machine learning for prediction in electronic health data. JAMA Netw Open 2018;1(4):e181404 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.1404] [Medline: 30646089]
24. Moehring RW, Phelan M, Lofgren E, Nelson A, Ashley ED, Anderson DJ, et al. Development of a machine learning model using electronic health record data to identify antibiotic use among hospitalized patients. JAMA Netw Open 2021;4(3):e213460 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.3460] [Medline: 33779743]
25. Ren Y, Loftus TJ, Datta S, Ruppert MM, Guan Z, Miao S, et al. Performance of a machine learning algorithm using electronic health record data to predict postoperative complications and report on a mobile platform. JAMA Netw Open 2022;5(5):e2211973 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.11973] [Medline: 35576007]
26. Yuan Q, Cai T, Hong C, Du M, Johnson BE, Lanuti M, et al. Performance of a machine learning algorithm using electronic health record data to identify and estimate survival in a longitudinal cohort of patients with lung cancer. JAMA Netw Open 2021;4(7):e2114723 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.14723] [Medline: 34232304]
27. Kamran F, Tang S, Otles E, McEvoy DS, Saleh SN, Gong J, et al. Early identification of patients admitted to hospital for COVID-19 at risk of clinical deterioration: model development and multisite external validation study. BMJ 2022;376:e068576 [FREE Full text] [doi: 10.1136/bmj-2021-068576] [Medline: 35177406]
28. Bennett TD, Moffitt RA, Hajagos JG, Amor B, Anand A, Bissell MM, National COVID Cohort Collaborative (N3C) Consortium. Clinical characterization and prediction of clinical severity of SARS-CoV-2 infection among US adults using data from the US national COVID cohort collaborative. JAMA Netw Open 2021;4(7):e2116901 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.16901] [Medline: 34255046]
29. Matheny ME, Ricket I, Goodrich CA, Shah RU, Stabler ME, Perkins AM, et al. Development of electronic health record-based prediction models for 30-day readmission risk among patients hospitalized for acute myocardial infarction. JAMA Netw Open 2021;4(1):e2035782 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.35782] [Medline: 33512518]
30. Liu K, Zhang X, Chen W, Yu ASL, Kellum JA, Matheny ME, et al. Development and validation of a personalized model with transfer learning for acute kidney injury risk estimation using electronic health records. JAMA Netw Open 2022;5(7):e2219776 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.19776] [Medline: 35796212]
31. Norgeot B, Glicksberg BS, Trupin L, Lituiev D, Gianfrancesco M, Oskotsky B, et al. Assessment of a deep learning model based on electronic health record data to forecast clinical outcomes in patients with rheumatoid arthritis. JAMA Netw Open 2019;2(3):e190606 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.0606] [Medline: 30874779]
32. Blecker S, Katz SD, Horwitz LI, Kuperman G, Park H, Gold A, et al. Comparison of approaches for heart failure case identification from electronic health record data. JAMA Cardiol 2016;1(9):1014-1020 [FREE Full text] [doi: 10.1001/jamacardio.2016.3236] [Medline: 27706470]
33. van der Ploeg T, Nieboer D, Steyerberg EW. Modern modeling techniques had limited external validity in predicting mortality from traumatic brain injury. J Clin Epidemiol 2016;78:83-89 [FREE Full text] [doi: 10.1016/j.jclinepi.2016.03.002] [Medline: 26987507]
34. Gravesteijn BY, Nieboer D, Ercole A, Lingsma HF, Nelson D, van Calster B, CENTER-TBI collaborators. Machine learning algorithms performed no better than regression models for prognostication in traumatic brain injury. J Clin Epidemiol 2020;122:95-107 [FREE Full text] [doi: 10.1016/j.jclinepi.2020.03.005] [Medline: 32201256]
35. Austin PC, Harrell FE, Steyerberg EW. Predictive performance of machine and statistical learning methods: impact of data-generating processes on external validity in the "large N, small p" setting. Stat Methods Med Res 2021;30(6):1465-1483 [FREE Full text] [doi: 10.1177/09622802211002867] [Medline: 33848231]
36. Wolpert DH, Macready WG. No free lunch theorems for optimization. IEEE Trans Evol Computat 1997;1(1):67-82. [doi: 10.1109/4235.585893]
37. Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015;162(1):W1-W73 [FREE Full text] [doi: 10.7326/M14-0698] [Medline: 25560730]
38. Inouye SK, van Dyck CH, Alessi CA, Balkin S, Siegal AP, Horwitz RI. Clarifying confusion: the confusion assessment method. A new method for detection of delirium. Ann Intern Med 1990;113(12):941-948. [doi: 10.7326/0003-4819-113-12-941] [Medline: 2240918]
39. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. New York, NY, United States: Association for Computing Machinery; 2016 Presented at: KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 13-17, 2016; San Francisco, California, USA p. 785-794 URL: https://dl.acm.org/doi/proceedings/10.1145/2939672 [doi: 10.1145/2939672.2939785]

XSL•FO
RenderX

40. Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc B 2018;58(1):267-288. [doi: 10.1111/j.2517-6161.1996.tb02080.x]
41. Stekhoven DJ, Bühlmann P. MissForest—non-parametric missing value imputation for mixed-type data. Bioinformatics 2012;28(1):112-118 [FREE Full text] [doi: 10.1093/bioinformatics/btr597] [Medline: 22039212]
42. Austin PC, van Klaveren D, Vergouwe Y, Nieboer D, Lee DS, Steyerberg EW. Validation of prediction models: examining temporal and geographic stability of baseline risk and estimated covariate effects. Diagn Progn Res 2017;1:12 [FREE Full text] [doi: 10.1186/s41512-017-0012-3] [Medline: 29350215]
43. Steyerburg EW. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. 2nd Edition. Switzerland: Springer Nature; 2019.
44. Matthews BW. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochim Biophys Acta 1975;405(2):442-451. [doi: 10.1016/0005-2795(75)90109-9] [Medline: 1180967]
45. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. 57 Morehouse Lane, Red Hook, NY, United States: Curran Associates Inc; 2017 Presented at: NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems; December 4-9, 2017; Long Beach California USA p. 4768-4777.
46. Nguyen TN, Haussen DC, Qureshi MM, Yamagami H, Fujinaka T, Mansour OY, SVIN COVID-19 Registry, the Middle East North Africa Stroke and Interventional Neurotherapies Organization (MENA-SINO), Japanese Society of Vascular and Interventional Neurology Society (JVIN). Decline in subarachnoid haemorrhage volumes associated with the first wave of the COVID-19 pandemic. Stroke Vasc Neurol 2021;6(4):542-552 [FREE Full text] [doi: 10.1136/svn-2020-000695] [Medline: 33771936]
47. Nogueira RG, Qureshi MM, Abdalkader M, Martins SO, Yamagami H, Qiu Z, SVIN COVID-19 Global Stroke Registry. Global impact of COVID-19 on stroke care and IV thrombolysis. Neurology 2021;96(23):e2824-e2838 [FREE Full text] [doi: 10.1212/WNL.0000000000011885] [Medline: 33766997]
48. Muckley ES, Saal JE, Meredig B, Roper CS, Martin JH. Interpretable models for extrapolation in scientific machine learning. ArXiv Preprint posted online on 16 December 2022. [FREE Full text] [doi: 10.1039/d3dd00082f]
49. Niculescu-Mizil A, Caruana R. Predicting good probabilities with supervised learning. New York, NY, United States: Association for Computing Machinery; 2005 Presented at: ICML '05: Proceedings of the 22nd International Conference on Machine Learning; August 7-11, 2005; Bonn Germany p. 625-632 URL: https://dl.acm.org/doi/proceedings/10.1145/1102351 [doi: 10.1145/1102351.1102430]

Abbreviations

AUPRC: area under the precision-recall curve
AUROC: area under the receiver operating characteristic curve
CAM: confusion assessment method
LASSO: least absolute shrinkage and selection operator
MCC: Matthews correlation coefficient
MHLW: Ministry of Health, Labour and Welfare
SHAP: Shapley additive explanations
XGBoost: extreme gradient boosting

Edited by T de Azevedo Cardoso; submitted 16.07.23; peer-reviewed by Y Ren, P Han; comments to author 05.09.23; revised version received 24.09.23; accepted 29.09.23; published 26.10.23.

Please cite as:
Matsumoto K, Nohara Y, Sakaguchi M, Takayama Y, Fukushige S, Soejima H, Nakashima N, Kamouchi M
Temporal Generalizability of Machine Learning Models for Predicting Postoperative Delirium Using Electronic Health Record Data: Model Development and Validation Study
JMIR Perioper Med 2023;6:e50895
URL: https://periop.jmir.org/2023/1/e50895
doi:10.2196/50895
PMID:37883164

©Koutarou Matsumoto, Yasunobu Nohara, Mikako Sakaguchi, Yohei Takayama, Syota Fukushige, Hidehisa Soejima, Naoki Nakashima, Masahiro Kamouchi. Originally published in JMIR Perioperative Medicine (http://periop.jmir.org), 26.10.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Perioperative Medicine, is properly cited. The complete bibliographic

XSL•FO
RenderX

information, a link to the original publication on http://periop.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A New Index for the Quantitative Evaluation of Surgical Invasiveness Based on Perioperative Patients' Behavior Patterns: Machine Learning Approach Using Triaxial Acceleration

Kozo Nakanishi[1], MD, PhD; Hidenori Goto[1], MD

Department of General Thoracic Surgery, National Hospital Organization Saitama Hospital, Wako Saitama, Japan

Corresponding Author:
Kozo Nakanishi, MD, PhD
Department of General Thoracic Surgery
National Hospital Organization Saitama Hospital
2-1 Suwa
Wako Saitama, 351-0102
Japan
Phone: 81 48 462 1101
Fax: 81 48 464 1138
Email: nakanishi.kozo.tf@mail.hosp.go.jp

Abstract

Background: The minimally invasive nature of thoracoscopic surgery is well recognized; however, the absence of a reliable evaluation method remains challenging. We hypothesized that the postoperative recovery speed is closely linked to surgical invasiveness, where recovery signifies the patient's behavior transition back to their preoperative state during the perioperative period.

Objective: This study aims to determine whether machine learning using triaxial acceleration data can effectively capture perioperative behavior changes and establish a quantitative index for quantifying variations in surgical invasiveness.

Methods: We trained 7 distinct machine learning models using a publicly available human acceleration data set as supervised data. The 3 top-performing models were selected to predict patient actions, as determined by the Matthews correlation coefficient scores. Two patients who underwent different levels of invasive thoracoscopic surgery were selected as participants. Acceleration data were collected via chest sensors for 8 hours during the preoperative and postoperative hospitalization days. These data were categorized into 4 actions (walking, standing, sitting, and lying down) using the selected models. The actions predicted by the model with intermediate results were adopted as the actions of the participants. The daily appearance probability was calculated for each action. The 2 differences between 2 appearance probabilities (sitting vs standing and lying down vs walking) were calculated using 2 coordinates on the x- and y-axes. A 2D vector composed of coordinate values was defined as the index of behavior pattern (iBP) for the day. All daily iBPs were graphed, and the enclosed area and distance between points were calculated and compared between participants to assess the relationship between changes in the indices and invasiveness.

Results: Patients 1 and 2 underwent lung lobectomy and incisional tumor biopsy, respectively. The selected predictive model was a light-gradient boosting model (mean Matthews correlation coefficient 0.98, SD 0.0027; accuracy: 0.98). The acceleration data yielded 548,466 points for patient 1 and 466,407 points for patient 2. The iBPs of patient 1 were [(0.32, 0.19), (–0.098, 0.46), (–0.15, 0.13), (–0.049, 0.22)] and those of patient 2 were [(0.55, 0.30), (0.77, 0.21), (0.60, 0.25), (0.61, 0.31)]. The enclosed areas were 0.077 and 0.0036 for patients 1 and 2, respectively. Notably, the distances for patient 1 were greater than those for patient 2 ({0.44, 0.46, 0.37, 0.26} vs {0.23, 0.0065, 0.059}; P=.03 [Mann-Whitney U test]).

Conclusions: The selected machine learning model effectively predicted the actions of the surgical patients with high accuracy. The temporal distribution of action times revealed changes in behavior patterns during the perioperative phase. The proposed index may facilitate the recognition and visualization of perioperative changes in patients and differences in surgical invasiveness.

(JMIR Perioper Med 2023;6:e50188) doi:10.2196/50188

KEYWORDS

surgery; invasiveness; triaxial acceleration; machine learning; human activity recognition; patient-oriented outcome; video-assisted thoracoscopic surgery; VATS; postoperative recovery; perioperative management; artificial intelligence; AI; mobile phone

XSL•FO
RenderX

Introduction

Challenges in Defining Invasiveness

Contemporary surgery favors minimally invasive techniques, synonymous with endoscopic procedures, which are now widely used. However, robust evidence of their minimal invasiveness is lacking [1]. This may be attributed to the absence of a reliable method for evaluating surgical invasiveness.

Despite the advent of modern endoscopic surgery and its various indicators, the definition of invasiveness remains controversial. Parameters are mostly derived from medical perspectives, such as adverse event rates, analgesic use, and hospital stay duration [2-9]. Nevertheless, statistical differences in such indicators, unrelated to the patients' daily lives, lack practicality and relevance [10].

Focus and Objective of the Study

On the basis of our postoperative observations, the rapid behavior recovery following endoscopic surgery surprised us, particularly during the early perioperative phase. This observation led us to believe that minimal invasiveness is associated with the swift restoration of patients' activity patterns.

Therefore, we aimed to establish an indicator that focuses on postoperative behavior recovery as a measure of surgical invasiveness. Patients gradually transition from rest to sitting, standing, walking, and finally returning to their preoperative lifestyles. A higher degree of invasiveness requires longer rest periods. This behavior change mirrors invasiveness. Knowing the behavior change as a quantitative indicator allows us to gauge invasiveness.

Although inpatient behavior patterns can be discerned by directly observing actions [11], human resources and privacy concerns arise, resulting in invasions of privacy. Therefore, we leveraged human activity recognition (HAR) technology to mitigate these issues by using machine learning and sensor data for activity detection [12-16].

Our investigation focused on the feasibility of understanding the behavior patterns of perioperative patients by using machine learning and a compact triaxial acceleration sensor. Moreover, we explored whether the knowledge of perioperative behavior patterns can yield a novel quantitative invasiveness index.

Methods

Ethical Considerations

This study was approved by the institutional ethics committee (R201519) and registered with the University Hospital Medical Information Network Individual Case Data Repository (UMIN000026843). We obtained written informed consent from all patients.

Patients

Overview

Among patients undergoing thoracoscopic surgery for lung lobectomy or tumor biopsy for thoracic malignancy at our institute (National Hospital Organization Saitama Hospital), 2 patients who could walk independently before surgery and who provided written consent to participate in this study were included. One patient was admitted to a 36 m^2 inpatient room with a toilet shared by 4 patients. The patient ate food from a bedside table and chair.

Perioperative medical care was provided in accordance with institutional clinical pathways. The patient was allowed to walk from the morning of postoperative day (POD) 1 with no obligatory transfers from the hospital room, except for daily chest radiographs. No behavior restrictions were imposed on the patient, and no rehabilitation was performed. Epidural analgesia was administered to control postoperative pain.

The following factors affecting the perioperative course were recorded: background (sex, age, and medical history); surgical procedure; operative time; blood loss; date of their walking resumption; drain removal date; amount of analgesics used; and duration of postoperative hospitalization.

A wearable sensor (myBeat; UNION-TOOL Corporation) 40.8×37.0×8.9 mm in size was used to measure accelerations, including gravity, in 3 orthogonal axes. The sensor was affixed to the center of the anterior chest of each participant. Data were measured from 9 AM to 5 PM on the preoperative day and during the postoperative period from POD 1 to the day before discharge.

Computer programs for data processing, analysis, and statistical testing were implemented using Python (version 3.9.16).

Selecting Learning Models to Predict Patients' Actions Among Classifiers

Seven classifier models were trained and evaluated: decision tree, logistic regression, linear-type support vector machine, kernel-type support vector machine, random forest (RF), gradient boosting classifier (GBC), and light-gradient boosting method (LGBM).

The classification models were trained using supervised data and the k-fold cross-validation method and then compared using the Matthews correlation coefficient (MCC) and accuracy as evaluation indices. A grid search was performed to determine optimal parameters. The features were validated using the correlation coefficient and importance scores among the decision tree–type estimators.

The 3 top-performing learned models with the highest MCC percentages were selected as the learning models to predict the patient's actions.

An open data set from the University of California School of Information and Computer Science repository was used as supervised data for machine learning. The data set was created at the Universitat Politècnica de Cataluña (UPC). Six actions were labeled in the UPC data set: WALKING, WALKING_UPSTAIRS, WALKING_DOWNSTAIRS, SITTING, STANDING, and LAYING. The data set contained 941,056 training data points. The raw acceleration training data from the data set were used as the supervised data set.

XSL•FO
RenderX

Data Preprocessing

Acceleration data were obtained as 3D vector data for the x-, y-, and z-axes. The axis directions were set to the x-axis of the data for gravity, y-axis for the patient's right side, and z-axis for the patient's rear side.

The measured data were separated into body acceleration vectors (BodyAcc) and gravity acceleration vectors (GravAcc) using bandpass filters. The BodyAcc magnitude was calculated as the root sum of the squares of BodyAcc. The Euler angles were calculated using GravAcc: the pitch and roll angles of the body axis in the direction of gravity.

The primary data set for calculating the features comprised 3 BodyAcc components, 3 GravAcc components, the BodyAcc magnitude, and 2 Euler angles.

The added features of the frequency component calculated from the metrics in the primary data set with fast Fourier transformation and 6 features for machine learning were used as test data: the median and SD of the acceleration value, frequency with the maximum amplitude, maximum amplitude value, phase, and mean frequency.

The primary data set and features for the training data were created from the raw data in the UPC data set using the same process as that used for the test data.

Prediction Aggregation

The actions of the patients were predicted using the 3 learning models selected. The actions predicted by the model that showed an intermediate prediction among the 3 models were considered as the actions of the patient.

In addition, the daily frequency of each activity was calculated. By dividing the frequency by the measured time of day, the proportion of time each action appeared per unit of time was calculated and defined as the action-specific indicator of the day, representing the appearance probability (AP). The appearance time (min) of an action per hour was calculated by multiplying the AP by 60.

To prevent accidents, patients were not allowed to use the stairs at the facility. Therefore, by combining 2 actions (WALKING_UPSTAIRS and _DOWNSTAIRS) with the WALKING action, 6 action classification labels were recounted into 4 categories: walking, standing, sitting, and lying. The 4 APs—walking (APwk), standing (APst), sitting (APsi), and lying down (APly)—were summed to 1.

Calculating Behavior Pattern Indices

A chart with x- and y-axes was prepared, and 4 vectors with 4 orthogonal directions were placed on the chart. The 4 vectors were APly at coordinates (0, APly); vAPwk (0, –APwk); vAPsi (APsi, 0); and vAPst (–APst, 0).

The index of behavior pattern (iBP) was created from the 4 AP vectors (*vAP*) to easily recognize the day-by-day transition of the AP balance in the chart. The new index was defined as the center-of-gravity vector coordinates of the 4 *vAP*. The iBP has a vector with coordinates (x, y) using the following formula:

$iBP(x, y) = ((APsi - APst) / 2, (APly - APwk) / 2)$ **(1)**

The *iBP* on each measurement day was plotted on a chart to evaluate the participants' iBP changes.

The following additional indicators were calculated to evaluate the iBP:

1. The area enclosed by the line segments sequentially connects the coordinates of the daily iBP.
2. Day-by-day distance between 2 sequential points of iBP from day-n to day-n+1.

$|iBP_{n+1} - iBP_n|$ (n=0, preoperatively)

1. Distance from the starting point: between the iBP point on day n and the preoperative day.

$|iBP_n - iBP_0|$ (n=0, preoperatively)

1. Total sum of the day-by-day distances from the starting point.

$\sum|iBP_n - iBP_0|$

Results

Patients

Table 1 presents the participants' characteristics. Patient 1 was a man aged 70 years who underwent thoracoscopic resection of the right lower lobe of the lung for lung cancer. Patient 2 was a man aged 71 years who had undergone thoracoscopic biopsy of an enlarged mediastinal mass (malignant lymphoma).

XSL•FO
RenderX

Table 1. Characteristics of the patients.

	Patient 1	Patient 2
Age (years)	70	71
Sex	Male	Male
Disease	Lung cancer	Mediastinal tumor (malignant lymphoma)
Surgical procedures	Right lower lobectomy under VATS[a]	Incisional biopsy of a tumor under VATS
Operation time (h: min)	5:06	1:48
Blood loss (mL)	186	11
Analgesic		
Continuous epidural analgia	Fentanyl citrate 0.72 mg mount	None, due to side effect
Other	Not used	Suppository (2 counts)
Postoperative course		
Resuming walk	POD[b] 1	POD 1
Removing the chest drain	POD 2	POD 2
Measured data		
Measurement period	Pre[c], POD 1-4	Pre, POD 1-3
Total number (in whole days)		
Measured acceleration data	548,466	466,407
All metric data	1,645,398	1,399,221

[a]VATS: video-assisted thoracoscopic surgery.

[b]POD: postoperative day.

[c]Pre: preoperative day.

Preprocessing Data

The total number of data points measured for the 2 participants was 1,014,873 (Table 1).

The measured data were recognized as 3 small-amplitude and high-frequency time-series signals with spike noise (Figure 1 provides an overview of the measured data of patient 1).

BodyAcc and GravAcc were extracted from the measured acceleration data using a high-pass filter with a stopband frequency of 1.5 Hz or a low-pass filter with a passband frequency of 3 Hz. BodyAcc was denoised using a low-pass filter with a stopband frequency of 10 Hz and a Hampel-type filter. GravAcc was denoised using a Hampel-type filter, and the moving average method was used at intervals of 50 (Figure 2 provides an overview of the denoised data of patient 1).

The Euler angles, which are the pitch and roll angles, were calculated from the extracted gravitational acceleration vectors with the direction of gravity as the axis (Figure 3). The pitch angle exhibits a characteristic stepwise waveform that reflects the movement of the upper body.

The labeled training data showed that 3 WALKING actions (WALKING, WALKING_UPSTAIRS, and WALKING_DOWNSTAIRS) were characterized by periodic changes in acceleration, whereas the acceleration values of the other actions were constant (Figure 4).

XSL•FO
RenderX

Figure 1. All measured triaxial acceleration data of the patients. (A) Patient 1; (B) Patient 2.

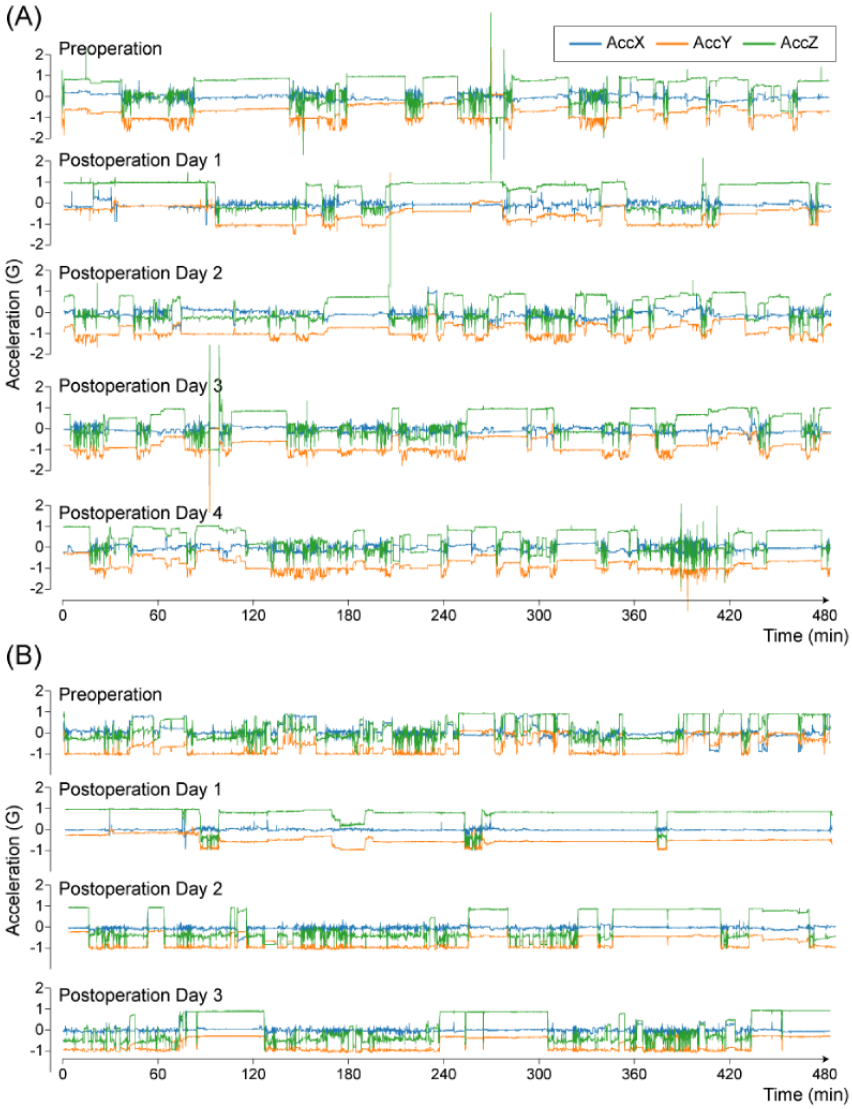


Figure 2. Body and gravitational acceleration data after noise elimination (data from patient 1 on the preoperative day).

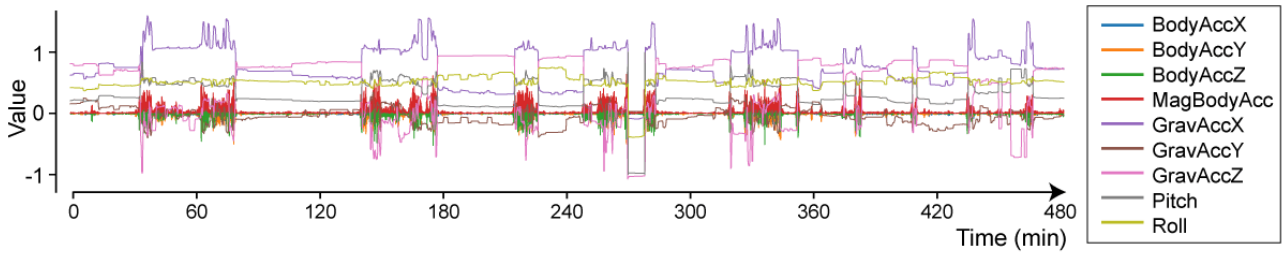
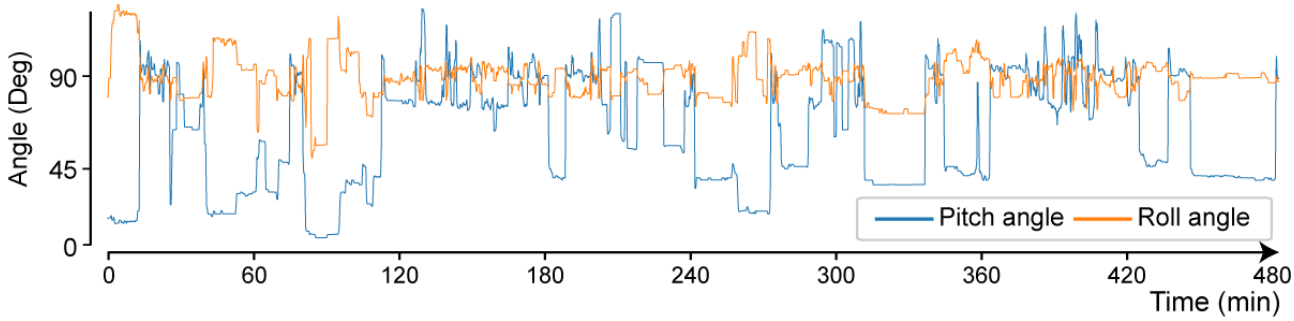
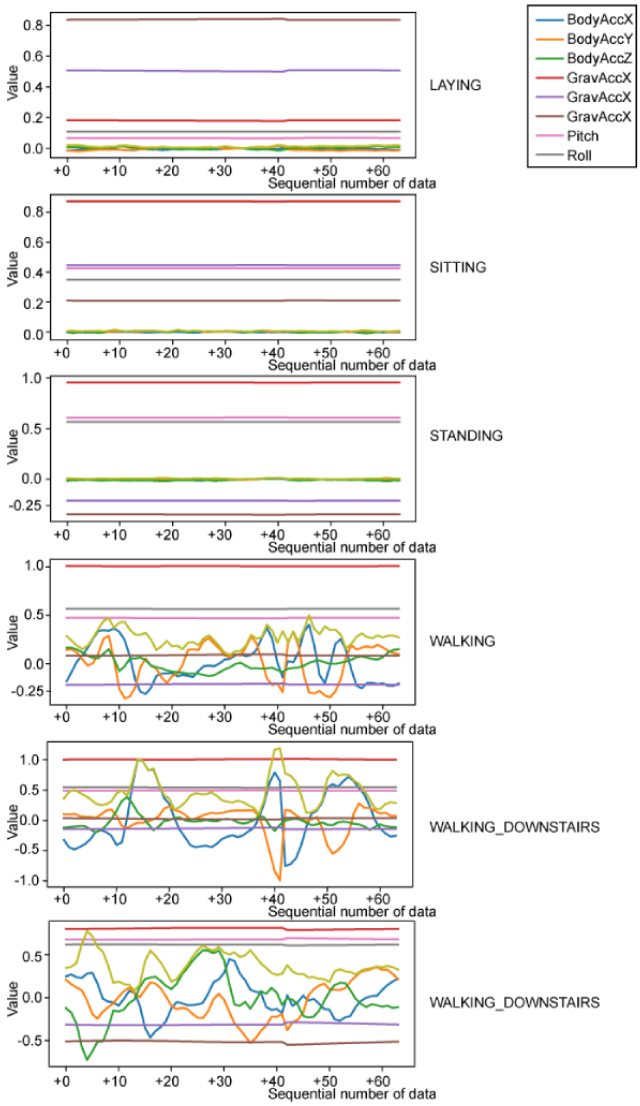


Figure 3. Pitch and roll angles (data from patient 1 on postoperative day 4).



XSL•FO
RenderX

Figure 4. Training data by labels (only part of the supervised data).



Selecting 3 Learning Models to Predict Actions

In validating the features, the correlation coefficients among the features were high between the median gpitch and yaw and between the "median Gravity Y" and "median Gravity Z" (Figure 5). However, the scatter graph of their correlations shows that they can be used to separate LAYING (Figure 6). Feature importance scores trended to be higher in "median Gravity X" and "median Gravity Y" and lower in features related to the frequency component. However, all features contributed to the prediction (Figure 7).

The cross-validation results for the trained models at the best-tuned parameters showed that the LGBM, GBC, and RF classifiers were the top performers with the highest mean MCCs, and the MCC values and accuracy were LGBM (mean MCC 0.98, SD 0.0027; mean accuracy: 0.98), GBC (mean 0.96, SD 0.0053; accuracy: 0.96), and RF (mean 0.95, SD 0.0079; accuracy: 0.95; Multimedia Appendix 1). The macroaverage accuracy scores for the best MCC scores were as follows: LGBM (MCC: 0.98; accuracy: 0.99; precision: 0.99; recall: 0.99; F_1-score: 0.99), GBC (MCC: 0.97; accuracy: 0.97; precision: 0.97; recall: 0.97; F_1-score: 0.97), and RF (MCC: 0.96; accuracy: 0.96; precision: 0.97; recall: 0.97; F_1-score: 0.97; Multimedia Appendix 2).

An analysis of the confusion matrices of the 3 classifiers showed that all had a 100% accuracy rate for LAYING, with no false positives or false negatives. Few prediction errors were observed between level walking and stair up or down, and most errors were between SITTING and STANDING (Figure 8).

XSL•FO
RenderX

Figure 5. Correlation heat map in terms of median feature data.

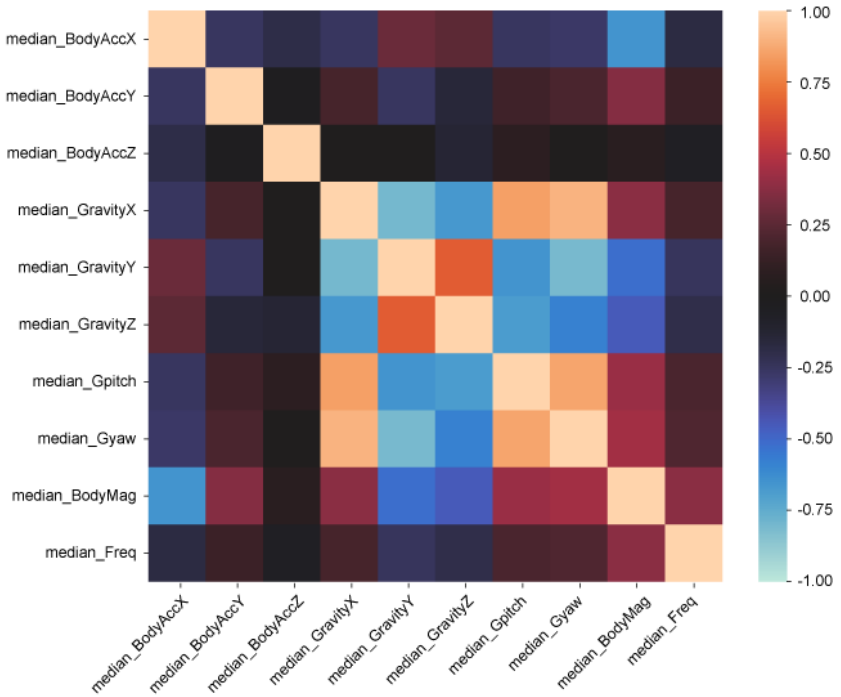
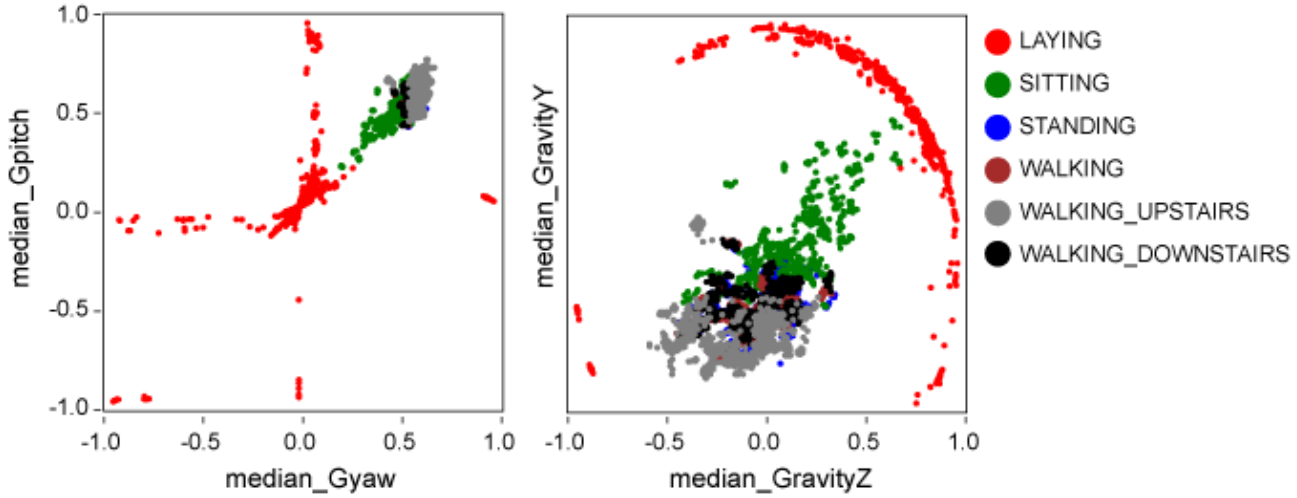


Figure 6. Correlations by label in features with high correlation coefficient.



XSL•FO
RenderX

Figure 7. Feature importance. GBC: gradient boosting classifier; DT: decision tree; RF: random forest; LGBM: light gradient boosting method.

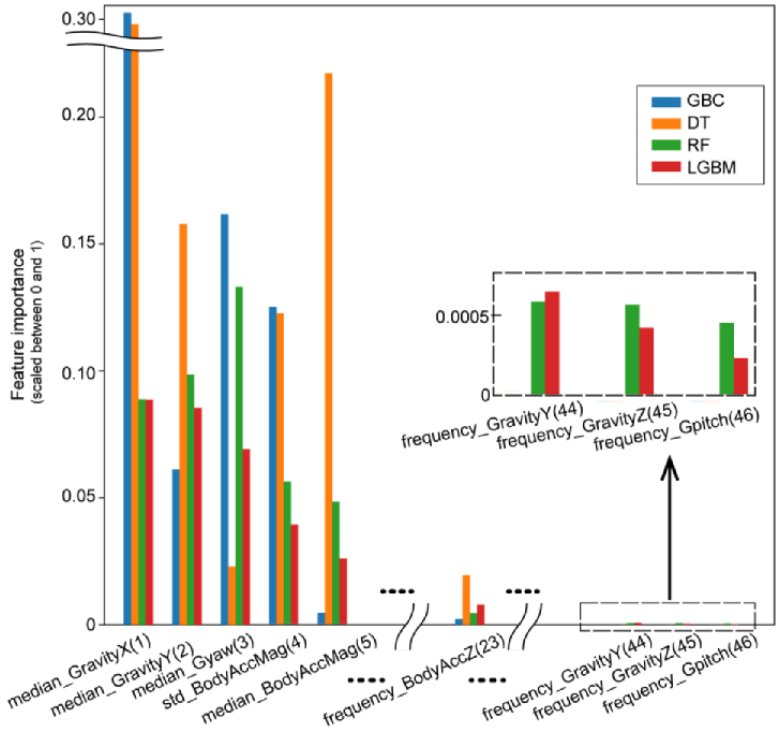
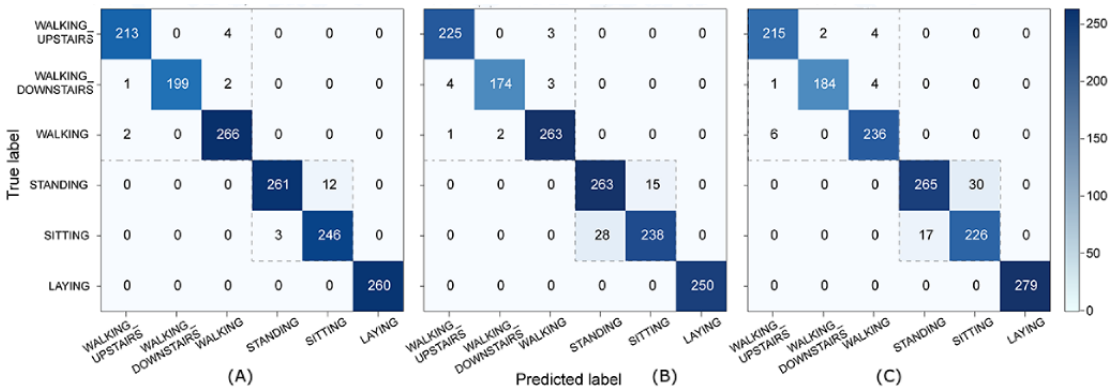


Figure 8. Confusion matrices of the best 3 learned classifiers. (A) Light gradient boosting method; (B) Gradient boosting classifier; (C) Random forest.



Aggregation of Predicted Actions

The 3 learned classifiers were provided with the test data and predicted actions. The appearance time per hour was calculated from the results (Multimedia Appendix 3).

Although slight differences in the prediction of the 3 classifiers are observed, they are generally consistent with the overall 8-hour measurement time (Figure 9).

The LGBM classifier exhibited intermediate values between the other 2 learned classifiers predicting actions using the test data. Therefore, the actions predicted by the LGBM classifier were considered the actions of the patients, and the action frequencies were calculated. Figure 10 shows the day-by-day changes in actions.

Patient 1 spent 11.5 minutes lying down per hour preoperatively and 0.4 minutes walking. On POD 1, the lying time increased to 27.5 minutes, and the walking time was 0.0 minutes. On POD 4, the lying down time decreased to 6.7 minutes, and the walking time increased to 0.2 minutes, which were near the preoperative values. The APly on POD 1 was significantly greater than that on the other days (P=.001; 1-sample 1-tailed t test).

Patient 2 had a walking time of 0.0 minutes per hour during hospitalization. Their standing time was 4.5 minutes preoperatively, which decreased to 0.5 minutes on POD 1; the lying time decreased from 17.9 to 12.7 minutes. Conversely, their sitting time increased from 37.7 to 44.0 minutes. On POD 2, the percentages of lying down, sitting, and standing times were close to the preoperative values. APsi on POD 1 was greater than that on other days (P=.01; 1-sample t test). APly on POD 1 was less than the preoperative value; however, the difference was statistically insignificant (P=.06; 1-sample t test).

The average APst of patient 1 was greater than that of patient 2 (P<.001; independent-sample t test), whereas APsi of patient 2 was greater than that of patient 1 (P=.005; independent-sample t test).

XSL•FO
RenderX

Figure 9. Action time diagram (patient 1, postoperative day 4). LGBM: light gradient boosting method. GBC: gradient boosting classifier. RF: random forest.

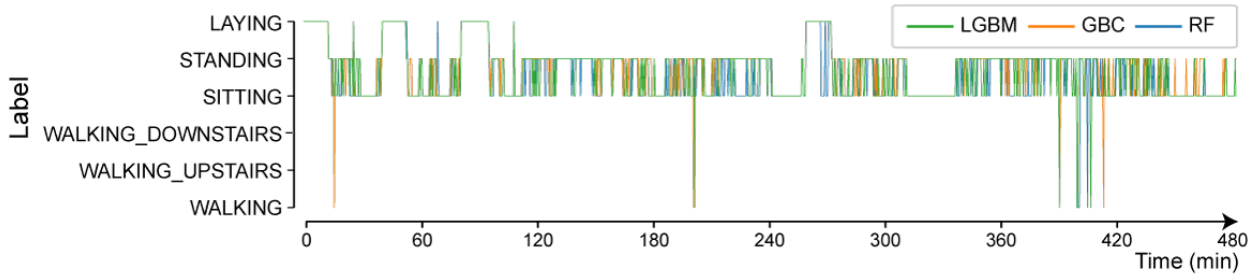
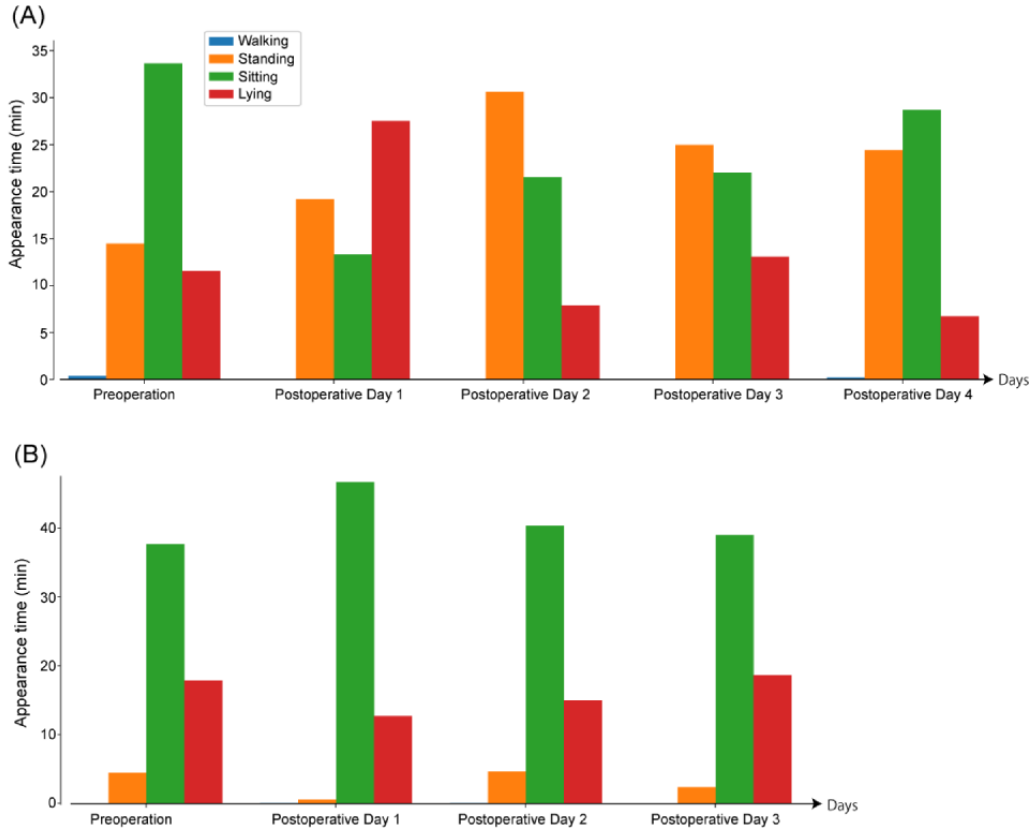


Figure 10. Day-by-day changes in the actions determined by the learned light-gradient boosting method classifier: (A) patient 1 and (B) patient 2.



Calculating Behavior Pattern Indices

The iBP values were calculated (Table 2) and plotted to determine the proportional changes for all actions (Figure 11).

In Figure 11, the preoperative index of the behavior pattern is expressed as point O and POD N is expressed as point N on the chart.

The regions enclosed by these points are colored. The area of the region bounded by the shift in coordinates was larger than that of patient 2. Figure 12 shows circles with equal areas enclosed by the iBP points. The differences between the 2 areas became clearer.

The accumulated bar on the left shows the sum of all distances, and the numbers indicate the distances. The bars on the right represent the distance from the preoperative day to the day of the procedure. “to POD N” means from preoperative day to POD N.

The distance gradually decreased after surgery. The values of patient 1 were greater than those of patient 2 (P=.03; Mann-Whitney U test).

Figure 13 illustrates the Euclidean distances between the iBP points on the surgery and preoperative days.

XSL•FO
RenderX

Table 2. Change in behavior pattern indices by date.

	Index of behavior patten (x, y)
Patient 1	
Pre[a]	(0.32, 0.19)
POD[b] 1	(−0.098, 0.46)
POD 2	(−0.15, 0.13)
POD 3	(−0.049, 0.22)
POD 4	(0.071, 0.11)
Patient 2	
Pre	(0.55, 0.30)
POD 1	(0.77, 0.21)
POD 2	(0.60, 0.25)
POD 3	(0.61, 0.31)

[a]Pre: preoperative day.
[b]POD: postoperative day.

Figure 11. Index of behavior pattern (iBP) on a graph: (A) patient 1 and (B) patient 2.

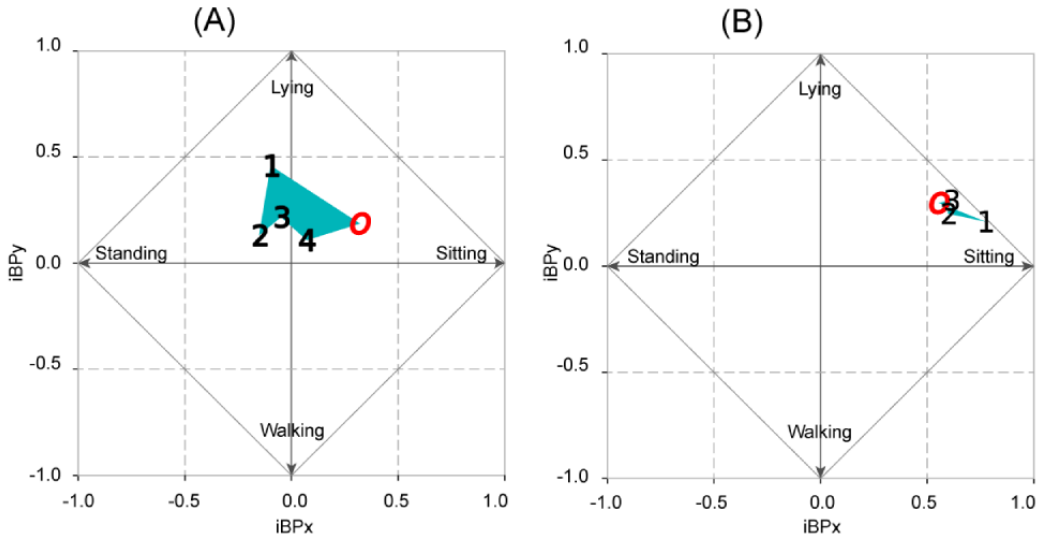
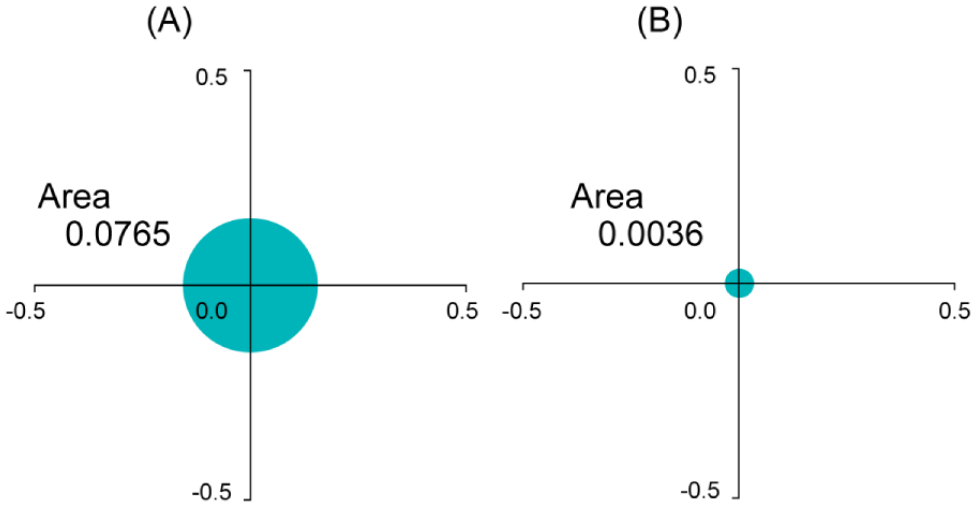
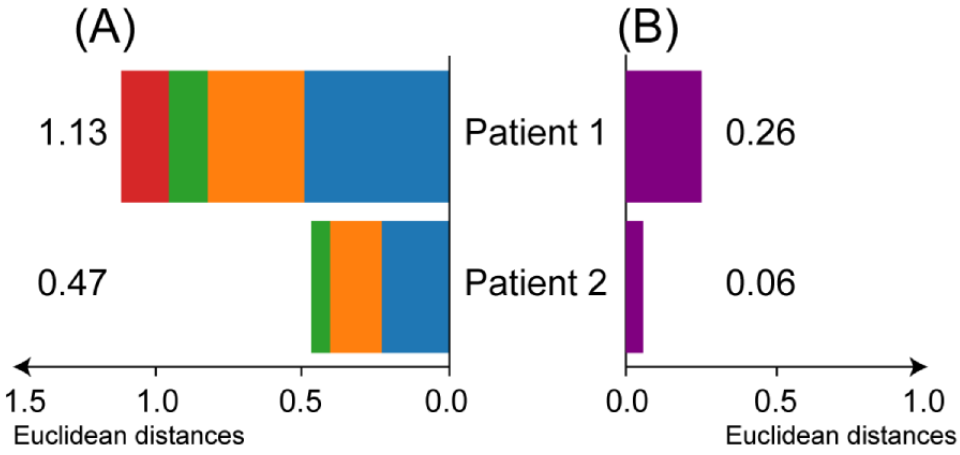


Figure 12. Differences in an area enclosed by behavior pattern indices: (A) patient 1 and (B) patient 2.



XSL•FO
RenderX

Figure 13. Euclidean distances between behavior pattern indices: (A) total distance and (B) last distance.



Discussion

Principal Findings

The principal findings are summarized as follows: first, by leveraging the latest artificial intelligence technology, the behavior patterns of perioperative patients can be accurately predicted using thoracic movement measurements alone. Second, a novel index derived from changes in behavior patterns demonstrated the potential to quantify surgical invasiveness.

The learned models (decision tree–type estimators) trained with an open data set accurately predicted the behavior actions of a surgical patient. The mean MCCs and accuracies were as follows: LGBM (0.98, 0.98), GBC (0.96, 0.96), and RF (0.95, 0.95). As shown in Figure 10, the time courses of actions predicted by the machine learning model based on the patient's acceleration data were aligned with the empirically known behavior changes in postoperative patients.

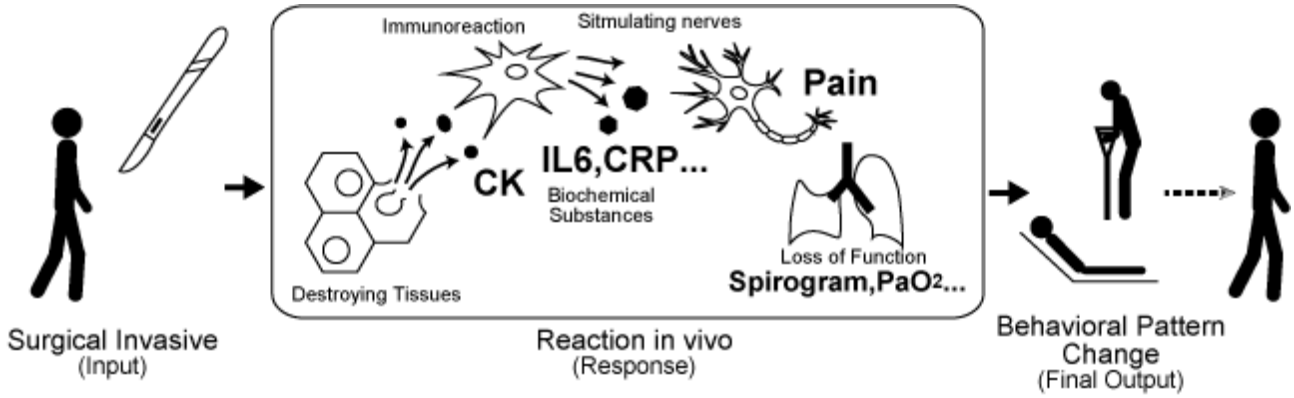
The proposed index, built on each patient's behavior patterns, effectively visualizes the disparities in physical activity among surgical patients with varying degrees of invasiveness. As shown in Figure 11, in both patients, point 1 (the index of POD 1) was farthest from point O (the preoperative index), and the points gradually moved closer to point O on each day. The differences in movements of the patients' indices were clearly visualized. The moving distances of patient 1 were significantly greater than those of patient 2 (P=.03; Mann-Whitney U test).

Comparison With Prior Work

New Perspective on Surgical Invasiveness

Compared with previous studies, a distinctive aspect of this research is the novel approach for assessing surgical invasiveness using the proposed index. This index was designed based on the final response to surgical invasion manifesting as changes in patients' behavior patterns during the early perioperative period (Figure 14). In addition, this study incorporated the latest advancements in machine learning technology.

Figure 14. Various indicators of surgical invasiveness. CK: creatine kinase. IL6: interleukin 6. CRP: C-reactive protein. PaO2: partial pressure of arterial oxygen.



Patient-Oriented Indicator Independent of Medical Parameters

Unlike conventional indices that rely on chemical or hospital-related parameters, the proposed index adopts a patient-oriented approach. It provides relevant and understandable information to patients, thereby facilitating their transition to daily life. As shown in Figure 11, point 2 of patient 2 has already returned to approximately point O. This indicates that on day 2, patient 2 could act similarly as before the surgery, without an additional load. The primary concern for patients undergoing invasive procedures is the time required to resume their preoperative lifestyle [17] and not the values of chemical or hospital indicators [7,8] unrelated to their daily life activities.

XSL•FO
RenderX

Perioperative Indicator Focusing on the Rest State

Although previous studies have often evaluated indicators several weeks after surgery [18-20], this study uniquely assessed the changes in behavior patterns during the early perioperative period. This distinction arises from this study, which focused on the remaining states of surgical patients. Although some prior studies have explored changes in behavior patterns after invasive events, they focused on active levels rather than activity at rest [21,22]. Numerous investigations have aimed to measure postoperative recovery using activity meters and step counts [23-27]; however, these efforts have yielded limited success, particularly during the perioperative period [14]. This might be attributed to the relatively short walking duration observed in hospitalized patients, both before and after the surgery [28-30]. The inherent limitations of the hospital environment, such as confined spaces, can restrict the patients' ability to walk extensively, diminishing the use of traditional methods to capture subtle variations in their daily activities [31].

We noted a change in the sitting time of patient 2 on POD 1. As expected, the lying time for patient 1 was prolonged after their major surgery. However, the sitting time for patient 2 was prolonged instead of their lying time. This may mean that the level of invasiveness experienced by patient 2 was insufficiently significant to require lying down but did require rest while sitting.

Vectorized Index

The proposed index is represented by vector data. Previous studies used simple scalar values [1,7-9,20,32]. Using only scalar indicators, comparisons between days and patients can be executed; vectorization, as illustrated in Figure 11, offers distinct advantages. Although 4D tensorization by the 4 APs is required to retain complete information, this complicates the visualization process.

Therefore, we transformed each scalar AP value into an AP vector by assigning directional values. Subsequently, to unify the 4 *AP* vectors into a single vector, we defined the center-of-gravity vector coordinates based on the 4 *AP* vectors. This vector was established as a new index, iBP.

When using a vectorized representation, the index captures both the direction and magnitude, both of which have clinical significance. By interpreting the preoperative index values as the patient's origin point, a vector directed toward this origin signified daily recovery. In both patients, the iBP returned was close to its origin, as expected.

The magnitude of the vector, which corresponds to the distance from the origin, reflects the magnitude of the patient's response to the surgical intervention. Any vector between the 2 indices denotes the magnitude of behavior change during that period, highlighting any increase or decrease in invasiveness.

Noninvasive Measurement Method During Perioperative Period

Acceleration was measured noninvasively, devoid of interventions that could impact the patients' recovery processes, such as exercise testing [33]. The small sensor affixed to the chest wall did not disrupt patient behavior during the perioperative period. This approach allows for painless acceleration measurement and enhances patient comfort and compliance. The sensor was removed at night during sleep to ensure the patient's ease.

Using the HAR Technology Clinical Research Method

HAR technology is a modern approach for monitoring and comprehending human movement and behavior by integrating sensor devices and algorithms, including machine learning [34]. Initially, HAR relied primarily on accelerometers [35], as used in this study. With subsequent advancements in sensor technology, it has evolved to encompass sensor fusion techniques that integrate various sensors, such as gyroscopes, magnetometers, and barometers, and physiological sensors, such as heart rate monitors. This integration enables real-time analyses of complex motions [36]. This technology has already been integrated into numerous smartphones and applied in diverse fields including drone and robot attitude controls [37]. In the medical field, particularly rehabilitation, analyses of walking strides [38] and fall-related events in patients [16] have been reported. The implementation of machine learning continues to emerge in the context of medical research. The HAR technology used in this study represents a fundamental aspect of the current HAR technology. Few studies, such as ours, have applied it as a clinical research method, underscoring its potential and significance [34,35].

Limitations

Validity of the New Index as an Invasiveness Indicator

This study did not assess the validity of the new index as an invasiveness indicator because of the absence of reliable methods for surgical invasiveness [1]. Consequently, we selected patients for 2 cases in which all surgeons agreed on the difference in surgical invasiveness. Although the 2 patients were of the same sex and similar in age, a notable disparity in surgical invasiveness existed between incisional biopsy and radical resection of the lung lobe and lymph nodes. The operation times were 306 and 108 minutes, and the blood loss volumes were 186 and 11 mL, respectively. In future studies, we plan to conduct further research using a larger number of cases.

Validity of the Categorization of Actions

We assumed that the breakdown of static position time was essential for evaluating recovery in the early postoperative period. Therefore, this study categorized patient behaviors into 3 static postures (ie, lying, sitting, and standing) and 1 dynamic action (ie, walking). The results sufficiently suggest the patient's behavior pattern for 8 hours during the daytime, although many other actions and postures are possible.

Current HAR technology cannot completely recognize quick dynamic actions or complex postures without periodicity [39]. Human postural changes are very complex, rapid, and varied to be accurately determined using machine learning. Considering the various actions for behavior prediction would result in poorer accuracy. Fortunately, we determined that walking times in hospitalized patients were short and that patients spent much time at rest. We assumed that other actions and behaviors accounted for a small part of the total action time during

XSL•FO
RenderX

hospitalization and could be ignored, such as changes in posture and elevator movements.

Validity of the Methodology for Predicting Behavior

Among the many machine learning methods, supervised learning methods are considered more reliable for tasks with statistically strong characteristic data, as in this study [1,13,28,40,41]. As shown in our results, the prediction accuracy was high and reliable. Using the latest artificial intelligence technology enabled us to convert approximately 1 million acceleration data points into a single paired numerical index.

Validity of Supervised Data Set

This study used the UPC data set [42] as supervised data because of its reliability. However, most publicly available data sets for HAR [37,43] are problematic because many volunteers supplying their data are young. Supervised data in future studies, particularly in those analyzing dynamic actions, should be collected from older adult volunteers and patients, if possible.

Validity of Selection of Measured Items and Features

Although the selected features produced accurate predictions consistent with the objectives of the study, the importance scores for the features associated with the frequency components were lower for the decision tree–type classifier. As a preliminary experiment, multiple iterations were performed using various UPC data set features. Despite these efforts, the prediction accuracy for the 4 action categories did not improve as expected. Consequently, only elementary statistics (median and SD) were selected as features for the triaxial acceleration data. In addition, the amplitude, frequency, and phase values after the fast Fourier transform were selected for detecting walking actions with a wave-like data morphology. Composite acceleration and Euler angles complemented the main items and were expected to reflect the patient's center-of-gravity acceleration, with a larger mean magnitude indicating more active movement. Among the added features, the median yaw angle of gravity ranked second in the importance scores, just after both the median Gravity X and Y, followed by the SD and median of BodyAccMag (Figure 7). The reason for the overall low importance scores of the frequency components is that the behavior classification in this study focused on static states without considering complex dynamic actions or postural transformations. Therefore, future studies on feature selection are required.

Limitation of Detection Using Only a Sensor With Triaxial Acceleration Data

In this study, we selected a sensor that is certified for medical applications in Japan. Using a single sensor presents a challenge in distinguishing between sitting and standing positions. Although the sensor captured only the triaxial acceleration, its prediction accuracy was satisfactory for our study objectives. Fixing the sensor to the anterior chest wall maintains the sensor direction constant and simplifies the behavior prediction. Many public HAR data sets stem from unfixed smartphones that lack direction and location specificity. Another factor that resulted in high accuracy was the surgical patient context. The patients could walk independently throughout the study; however, their behavior was curtailed in the hospital, devoid of running or long walks. We attribute these factors to high prediction accuracy. Although attaching an additional sensor can potentially broaden the range of classified actions, multiple sensors introduce technical complexities such as maintaining a precise synchronization.

Future Directions

Future studies should validate the proposed index with more cases to determine whether it reflects invasiveness.

Nonetheless, the applicability of this new index is not limited to patients who have undergone surgery. Because it relies solely on patient behavior, it can be extended to evaluate nonsurgically hospitalized patients or individuals requiring nursing care. Daily behavior patterns were encapsulated in an index expressed as a numerical pair between 0 and 1, facilitating statistical comparisons. This inclusivity can extend beyond daily changes within the same patient for comparisons between different patients and procedures.

With the ongoing development of the HAR technology, detailed research on the behavior patterns of patients with more severe conditions or more active patients is possible. Currently, the sensors are equipped with more functions and are smaller than those when this study was designed. The device can detect differences of several centimeters in height based on barometric pressure [44] and track movements using geomagnetic and satellite data. We will be able to grasp the difference between a patient's sitting and standing positions more easily as well as their speed and range of movement.

We hope that this study inspires medical professionals to conduct medical research using machine learning and to apply the proposed index, conception, and methodology.

Conclusions

To establish a quantitative measurement method for surgical invasiveness, we focused on behavior patterns in operative patients and proposed an index for the rate of postural time. The index was created using a supervised machine learning method based on the triaxial acceleration data of a patient before and after surgery. The numerical indices clearly demonstrate the difference between 2 patients with different levels of surgical invasiveness on a graph and enable numerical comparisons. The proposed index was created using parameters that do not depend on the type of invasiveness, and we believe that it can be widely applied beyond the evaluation of surgical invasiveness.

Acknowledgments

The authors thank Davide Angenia, Alessandro Ghio, Luca Oneto, Xavier Parra, and Jorge L Reyes-Ortiz for providing the public domain data set, A Public Domain Dataset for Human Activity Recognition Using Smartphones.

XSL•FO
RenderX

Data Availability

The data are available upon reasonable request from the corresponding author.

Authors' Contributions

KN was responsible for conceptualization, methodology, software programming, project administration, writing, reviewing, and editing. HG collected the acceleration data, performed the surgeries, and wrote the original draft.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Matthews correlation coefficients of the classifiers.
[PDF File (Adobe PDF File), 97 KB - periop_v6i1e50188_app1.pdf]

Multimedia Appendix 2
Accuracy classification score among classifiers with the best Matthews correlation coefficients.
[PDF File (Adobe PDF File), 51 KB - periop_v6i1e50188_app2.pdf]

Multimedia Appendix 3
The change in the appearance time of the patient's actions was predicted by the 3 best learned classifiers.
[PDF File (Adobe PDF File), 211 KB - periop_v6i1e50188_app3.pdf]

References

1. Alam N, Flores RM. Video-assisted thoracic surgery (VATS) lobectomy: the evidence base. JSLS 2007 Jul;11(3):368-374 [FREE Full text] [Medline: 17931521]
2. Nagahiro I, Andou A, Aoe M, Sano Y, Date H, Shimizu N. Pulmonary function, postoperative pain, and serum cytokine level after lobectomy: a comparison of VATS and conventional procedure. Ann Thorac Surg 2001 Aug;72(2):362-365. [doi: 10.1016/s0003-4975(01)02804-1] [Medline: 11515867]
3. Paul S, Altorki NK, Sheng S, Lee PC, Harpole DH, Onaitis MW, et al. Thoracoscopic lobectomy is associated with lower morbidity than open lobectomy: a propensity-matched analysis from the STS database. J Thorac Cardiovasc Surg 2010 Feb;139(2):366-378 [FREE Full text] [doi: 10.1016/j.jtcvs.2009.08.026] [Medline: 20106398]
4. Nicastri DG, Wisnivesky JP, Litle VR, Yun J, Chin C, Dembitzer FR, et al. Thoracoscopic lobectomy: report on safety, discharge independence, pain, and chemotherapy tolerance. J Thorac Cardiovasc Surg 2008 Mar;135(3):642-647 [FREE Full text] [doi: 10.1016/j.jtcvs.2007.09.014] [Medline: 18329487]
5. Matsuura N, Igai H, Ohsawa F, Yazawa T, Kamiyoshihara M. Uniport multiport video-assisted thoracoscopic surgery for anatomical lung resection-which is less invasive? J Thorac Dis 2021 Jan;13(1):244-251 [FREE Full text] [doi: 10.21037/jtd-20-2759] [Medline: 33569204]
6. Landreneau RJ, Mack MJ, Hazelrigg SR, Naunheim K, Dowling RD, Ritter P, et al. Prevalence of chronic pain after pulmonary resection by thoracotomy or video-assisted thoracic surgery. J Thorac Cardiovasc Surg 1994 Apr;107(4):1079-1085 [FREE Full text] [doi: 10.1097/00132586-199412000-00051] [Medline: 8159030]
7. Craig SR, Leaver HA, Yap PL, Pugh GC, Walker WS. Acute phase responses following minimal access and conventional thoracic surgery. Eur J Cardiothorac Surg 2001 Sep;20(3):455-463. [doi: 10.1016/s1010-7940(01)00841-7] [Medline: 11509263]
8. Shigemura N, Akashi A, Nakagiri T, Ohta M, Matsuda H. Complete versus assisted thoracoscopic approach: a prospective randomized trial comparing a variety of video-assisted thoracoscopic lobectomy techniques. Surg Endosc 2004 Oct;18(10):1492-1497. [doi: 10.1007/s00464-003-8252-4] [Medline: 15791376]
9. Demmy TL, Curtis JJ. Minimally invasive lobectomy directed toward frail and high-risk patients: a case-control study. Ann Thorac Surg 1999 Jul;68(1):194-200. [doi: 10.1016/s0003-4975(99)00467-1] [Medline: 10421140]
10. Low CA, Bovbjerg DH, Ahrendt S, Choudry MH, Holtzman M, Jones HL, et al. Fitbit step counts during inpatient recovery from cancer surgery as a predictor of readmission. Ann Behav Med 2018 Jan 05;52(1):88-92 [FREE Full text] [doi: 10.1093/abm/kax022] [Medline: 29538623]
11. Lubans DR, Hesketh K, Cliff DP, Barnett LM, Salmon J, Dollman J, et al. A systematic review of the validity and reliability of sedentary behaviour measures used with children and adolescents. Obes Rev 2011 Oct;12(10):781-799. [doi: 10.1111/j.1467-789X.2011.00896.x] [Medline: 21676153]
12. Demrozi F, Pravadelli G, Bihorac A, Rashidi P. Human activity recognition using inertial, physiological and environmental sensors: a comprehensive survey. IEEE Access 2020;8:210816-210836 [FREE Full text] [doi: 10.1109/access.2020.3037715] [Medline: 33344100]

XSL•FO
RenderX

13. Chen K, Zhang D, Yao L, Guo B, Yu Z, Liu Y. Deep learning for sensor-based human activity recognition: overview, challenges, and opportunities. ACM Comput Surv 2021 May 24;54(4):1-40 [FREE Full text] [doi: 10.1145/3447744]
14. Broderick JM, Ryan J, O'Donnell DM, Hussey J. A guide to assessing physical activity using accelerometry in cancer patients. Support Care Cancer 2014 Apr;22(4):1121-1130. [doi: 10.1007/s00520-013-2102-2] [Medline: 24389829]
15. Yang X, Zhao Y, Street GM, Huang Y, Filip To SD, Purswell JL. Classification of broiler behaviours using triaxial accelerometer and machine learning. Animal 2021 Jul;15(7):100269 [FREE Full text] [doi: 10.1016/j.animal.2021.100269] [Medline: 34102430]
16. Chelli A, Patzold M. A machine learning approach for fall detection and daily living activity recognition. IEEE Access 2019;7:38670-38687 [FREE Full text] [doi: 10.1109/access.2019.2906693]
17. Montroni I, Ugolini G, Audisio RA. Functional recovery and patient-reported outcomes after surgery. Aging Clin Exp Res 2018 Mar;30(3):259-262. [doi: 10.1007/s40520-017-0867-8] [Medline: 29305794]
18. Carmichael H, Overbey DM, Hosokawa P, Goode CM, Jones TS, Barnett CCJ, et al. Wearable technology-a pilot study to define "normal" postoperative recovery trajectories. J Surg Res 2019 Dec;244:368-373. [doi: 10.1016/j.jss.2019.06.057] [Medline: 31323392]
19. Jonsson M, Ahlsson A, Hurtig-Wennlöf A, Vidlund M, Cao Y, Westerdahl E. In-hospital physiotherapy and physical recovery 3 months after lung cancer surgery: a randomized controlled trial. Integr Cancer Ther 2019 Jan;18:1534735419876346 [FREE Full text] [doi: 10.1177/1534735419876346] [Medline: 31530046]
20. Salati M, Brunelli A, Xiumè F, Monteverde M, Sabbatini A, Tiberi M, et al. Video-assisted thoracic surgery lobectomy does not offer any functional recovery advantage in comparison to the open approach 3 months after the operation: a case matched analysis†. Eur J Cardiothorac Surg 2017 Jun 01;51(6):1177-1182. [doi: 10.1093/ejcts/ezx013] [Medline: 28329201]
21. Moscarelli M, Lorusso R, Abdullahi Y, Varone E, Marotta M, Solinas M, et al. The effect of minimally invasive surgery and sternotomy on physical activity and quality of life. Heart Lung Circ 2021 Jun;30(6):882-887. [doi: 10.1016/j.hlc.2020.09.936] [Medline: 33191139]
22. Labib HS, Fawaz SI, Ghanam ME, ELBarbray MG. Effect of minimally invasive cardiac surgery compared with conventional surgery on post-operative physical activity and rehabilitation in patients with valvular heart disease. Egypt Rheumatol Rehabil 2023 Jan 26;50(1):5 [FREE Full text] [doi: 10.1186/s43166-023-00171-3]
23. Jonker LT, Hendriks S, Lahr MM, van Munster BC, de Bock GH, van Leeuwen BL. Postoperative recovery of accelerometer-based physical activity in older cancer patients. Eur J Surg Oncol 2020 Nov;46(11):2083-2090 [FREE Full text] [doi: 10.1016/j.ejso.2020.06.012] [Medline: 32682650]
24. van der Meij E, van der Ploeg HP, van den Heuvel B, Dwars BJ, Meijerink WJ, Bonjer HJ, et al. Assessing pre- and postoperative activity levels with an accelerometer: a proof of concept study. BMC Surg 2017 May 12;17(1):56 [FREE Full text] [doi: 10.1186/s12893-017-0223-0] [Medline: 28494785]
25. Iida H, Shimizu T, Maehira H, Kitamura N, Mori H, Miyake T, et al. A pilot study: the association between physical activity level using by accelerometer and postoperative complications after hepatic resection. Exp Ther Med 2018 Dec;16(6):4893-4899 [FREE Full text] [doi: 10.3892/etm.2018.6816] [Medline: 30542445]
26. Brown CH4, Yanek L, Healy R, Tsay T, Di J, Goeddel L, Cardiac Surgery Mobility Working Group. Comparing three wearable accelerometers to measure early activity after cardiac surgery. JTCVS Open 2022 May 31;11:176-191 [FREE Full text] [doi: 10.1016/j.xjon.2022.05.011] [Medline: 36172447]
27. Melanson ELJ, Freedson PS. Validity of the Computer Science and Applications, Inc. (CSA) activity monitor. Med Sci Sports Exerc 1995 Jun;27(6):934-940. [Medline: 7658958]
28. Browning L, Denehy L, Scholes RL. The quantity of early upright mobilisation performed following upper abdominal surgery is low: an observational study. Aust J Physiother 2007;53(1):47-52. [doi: 10.1016/s0004-9514(07)70061-2] [Medline: 17326738]
29. Agostini PJ, Naidu B, Rajesh P, Steyn R, Bishay E, Kalkat M, et al. Potentially modifiable factors contribute to limitation in physical activity following thoracotomy and lung resection: a prospective observational study. J Cardiothorac Surg 2014 Sep 27;9:128 [FREE Full text] [doi: 10.1186/1749-8090-9-128] [Medline: 25262229]
30. Fazio S, Stocking J, Kuhn B, Doroy A, Blackmon E, Young HM, et al. How much do hospitalized adults move? A systematic review and meta-analysis. Appl Nurs Res 2020 Feb;51:151189 [FREE Full text] [doi: 10.1016/j.apnr.2019.151189] [Medline: 31672262]
31. Kozey SL, Lyden K, Howe CA, Staudenmayer JW, Freedson PS. Accelerometer output and MET values of common physical activities. Med Sci Sports Exerc 2010 Sep;42(9):1776-1784 [FREE Full text] [doi: 10.1249/MSS.0b013e3181d479f2] [Medline: 20142781]
32. Begum SS, Papagiannopoulos K, Falcoz PE, Decaluwe H, Salati M, Brunelli A. Outcome after video-assisted thoracoscopic surgery and open pulmonary lobectomy in patients with low VO2 max: a case-matched analysis from the ESTS database†. Eur J Cardiothorac Surg 2016 Apr;49(4):1054-1058. [doi: 10.1093/ejcts/ezv378] [Medline: 26604295]
33. Nomori H, Horio H, Naruke T, Suemasu K. What is the advantage of a thoracoscopic lobectomy over a limited thoracotomy procedure for lung cancer surgery? Ann Thorac Surg 2001 Sep;72(3):879-884. [doi: 10.1016/s0003-4975(01)02891-0] [Medline: 11565674]

XSL•FO
RenderX

34. Yadav SK, Tiwari K, Pandey HM, Akbar SA. A review of multimodal human activity recognition with special emphasis on classification, applications, challenges and future directions. Knowl Based Syst 2021 Jul 08;223:106970 [FREE Full text] [doi: 10.1016/j.knosys.2021.106970]
35. Aboo AK, Ibrahim LM. Survey on human activity recognition using smartphone. Raf J Comput Sci Math 2021 Jun 01;15(1):55-67 [FREE Full text] [doi: 10.33899/csmj.2021.168253]
36. Wan S, Qi L, Xu X, Tong C, Gu Z. Deep learning models for real-time human activity recognition with smartphones. Mobile Netw Appl 2020;25(2):743-755 [FREE Full text] [doi: 10.1007/s11036-019-01445-x]
37. De-La-Hoz-Franco E, Ariza-Colpas P, Quero JM, Espinilla M. Sensor-based datasets for human activity recognition – a systematic review of literature. IEEE Access 2018;6:59192-59210 [FREE Full text] [doi: 10.1109/access.2018.2873502]
38. Trabassi D, Serrao M, Varrecchia T, Ranavolo A, Coppola G, De Icco R, et al. Machine learning approach to support the detection of Parkinson's disease in IMU-based gait analysis. Sensors (Basel) 2022 May 12;22(10):3700 [FREE Full text] [doi: 10.3390/s22103700] [Medline: 35632109]
39. Reyes-Ortiz JL, Oneto L, Samà A, Parra X, Anguita D. Transition-aware human activity recognition using smartphones. Neurocomputing 2016 Jan;171:754-767 [FREE Full text] [doi: 10.1016/j.neucom.2015.07.085]
40. McCarthy MW, James DA, Lee JB, Rowlands DD. Decision‐tree‐based human activity classification algorithm using single‐channel foot‐mounted gyroscope. Electron Lett 2015 Apr;51(9):675-676 [FREE Full text] [doi: 10.1049/el.2015.0436]
41. van Dijk-Huisman HC, Bijnens W, Senden R, Essers JM, Meijer K, Aarts J, et al. Optimization and validation of a classification algorithm for assessment of physical activity in hospitalized patients. Sensors (Basel) 2021 Feb 27;21(5):1652 [FREE Full text] [doi: 10.3390/s21051652] [Medline: 33673447]
42. Anguita D, Ghio A, Oneto L, Parra X, Reyes-Ortiz J. A public domain dataset for human activity recognition using smartphones. In: Proceedings of the 21st European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. 2013 Presented at: ESANN '13; April 24-26, 2013; Bruges, Belgium p. 437-434. [doi: 10.1007/978-3-642-40728-4_54]
43. Garcia-Gonzalez D, Rivero D, Fernandez-Blanco E, Luaces MR. A public domain dataset for real-life human activity recognition using smartphone sensors. Sensors (Basel) 2020 Apr 13;20(8):2200 [FREE Full text] [doi: 10.3390/s20082200] [Medline: 32295028]
44. DPS310 datasheet. Infineon Technologies. URL: https://www.infineon.com/dgdl/Infineon-DPS310-DataSheet-v01_02-EN.pdf?fileId=5546d462576f34750157750826c42242 [accessed 2023-09-12]

Abbreviations

AP: appearance probability
GBC: gradient boosting classifier
HAR: human activity recognition
iBP: index of behavior pattern
LGBM: light-gradient boosting method
MCC: Matthews correlation coefficient
POD: postoperative day
RF: random forest
UPC: Universitat Politècnica de Cataluña

Edited by T Leung, T de Azevedo Cardoso; submitted 22.06.23; peer-reviewed by PF Chen, D Trabassi; comments to author 07.08.23; revised version received 12.09.23; accepted 11.10.23; published 14.11.23.

Please cite as:
Nakanishi K, Goto H
A New Index for the Quantitative Evaluation of Surgical Invasiveness Based on Perioperative Patients' Behavior Patterns: Machine Learning Approach Using Triaxial Acceleration
JMIR Perioper Med 2023;6:e50188
URL: https://periop.jmir.org/2023/1/e50188
doi:10.2196/50188
PMID:37962919

©Kozo Nakanishi, Hidenori Goto. Originally published in JMIR Perioperative Medicine (http://periop.jmir.org), 14.11.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium,

XSL•FO
RenderX

provided the original work, first published in JMIR Perioperative Medicine, is properly cited. The complete bibliographic information, a link to the original publication on http://periop.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Editorial

JMIR Perioperative Medicine: A Global Journal for Publishing Interdisciplinary Innovations, Research, and Perspectives

Nidhi Rohatgi[1], MD, MS

Stanford University School of Medicine, Stanford, CA, United States

Corresponding Author:
Nidhi Rohatgi, MD, MS
Stanford University School of Medicine
300 Pasteur Drive
Stanford, CA, 94304
United States
Phone: 1 650 725 4890
Email: nrohatgi@stanford.edu

Abstract

JMIR Perioperative Medicine supports the dissemination of technological and data science–driven innovative research conducted by interdisciplinary teams in perioperative medicine. We invite contributions on a broad range of topics from clinicians, scientists, and allied health professionals from across the globe.

(JMIR Perioper Med 2023;6:e54344) doi:10.2196/54344

KEYWORDS

JMIR Perioperative Medicine; innovation; technology; digital health; research; interdisciplinary; perioperative medicine

Perioperative Medicine: From Opium to Artificial Intelligence

It was not until 1846 that the possibility of undergoing a major surgical procedure without experiencing excruciating pain was recorded [1]; this surgery was performed in the United States. Before this, it was opium, mandrake, alcohol, or simply "biting the bullet" for invasive procedures. This was also the first record of the use of technology in the operating theater with the introduction of the first "anesthetic machine" [2]. A few decades later, a Scottish surgeon introduced the concept of orotracheal intubation, and a century later, the modern ventilator was developed in Sweden [3]. All these discoveries and innovations led to seismic shifts in the field of perioperative medicine. As surgical and anesthetic techniques evolved, focus on systematic perioperative data recording and patient monitoring, outcomes, and experiences also emerged. In 1949, the idea of preoperative clinics was introduced in England "so the patients can arrive in the operating theater as strong and healthy as possible" [4]. In 2007, the World Health Organization created a surgical safety checklist that decreased surgical morbidity and mortality across 8 hospitals in different countries, and now, over 75% of institutions across 94 countries use these checklists [5]. In 2020, amid the COVID-19 pandemic, as surgeries came close to a standstill, exemplar collaborations across multiple high- and low-income countries occurred to study surgical safety in patients with COVID-19 [6]. Time and time again, the publication of an "aha moment," an incredible breakthrough, in one corner of the world catalyzed a change in the care of surgical patients and the practice of perioperative medicine across the globe.

Today, technological advancements and digital innovations continue to transform perioperative medicine at an unprecedented pace compared to yesteryears. This field has become much more interdisciplinary than ever before. There has been an increasing amount of research to predict, prevent, or understand the short- and long-term impacts of physiological perturbances during the perioperative period on different organ systems. mHealth (mobile health) technologies are being used to educate surgical patients, caregivers, and clinicians about perioperative pain management or to collect patient health data, perioperative outcomes, or experiences. Resource utilization, access, and patient and clinician experience are being optimized through telemedicine and digital communication strategies [7]. Big data is being leveraged to drive quality improvement and research in some centers while others are driving innovations in the midst of paper charts [8]. Biases in the use of common devices such as pulse oximeters are being recognized in North America while some countries continue to struggle to get access to these common devices [9]. Artificial intelligence and machine learning, which seemed quixotic years ago, have made an unexpectedly large leap with several potential applications, such as reducing medication errors, perioperative risk assessment, early detection of clinical deterioration, predicting postoperative

XSL•FO
RenderX

morbidity and mortality, streamlining workflows, creating personalized care pathways, real-time image analysis, providing clinical decision support, analysis of structured and unstructured data in electronic health records, or understanding health disparities [10-12]. Experts in perioperative medicine are investigating a broad range of topics—from sustainability and the carbon footprint of surgeries to taking a deeper dive into precision medicine using omics and biomarkers [13]. Some of these discoveries and research will reduce inequities and improve surgical safety on a global scale while others may improve outcomes and care delivery locally, both of which are important.

JMIR Perioperative Medicine: Focus and Scope

JMIR Perioperative Medicine is a global, peer-reviewed, open access journal indexed in PubMed, PubMed Central, Directory of Open Access Journals (DOAJ Seal), EBSCO/EBSCO Essentials, and Sherpa/Romeo. This journal welcomes contributions from a diverse group of investigators and clinicians from across the globe. We accept clinical trials, meta-analyses, reviews, observational studies, quality improvement studies, research letters, viewpoints, and tutorials. We are committed to disseminating research and advancements in all aspects of perioperative medicine. This includes novel and promising formative studies that may inform future research in the field. We accept articles on a broad range of topics in perioperative medicine such as digital and technological innovations; patient monitoring; patient and clinician education; operating room management; anesthetic techniques; pain management; interdisciplinary care; perioperative risk assessment; blood management; best practices; access to surgery and critical care; quality; safety; cost-effectiveness; diversity, equity, and inclusion; climate change and sustainability; and global surgical initiatives.

Although substantial progress has been made in perioperative medicine, much remains to be done and new frontiers are yet to be explored. Life expectancy has increased and older patients with many more comorbidities are opting to undergo surgical procedures. Globally, 4.2 million patients still die within 30 days of surgery, with over half of these deaths occurring in low- and middle-income countries [14]. This amounts to 7.7% of deaths globally, making postoperative deaths the third leading contributor to mortality in the world after ischemic heart disease and stroke. This is a staggering statistic. There remain several unresolved clinical conundrums in perioperative medicine. Health care delivery is also tasked with creating innovative solutions to tackle workflow optimization, cost-effectiveness, emergency preparedness, workforce wellness, education, and communication, all while achieving the highest-quality patient outcomes and experiences. Perioperative medicine continues to face an attractive challenge, one that motivates all of us to think creatively, share and publish our knowledge and experiences, and build on each other's ideas and innovations from across the world. *JMIR Perioperative Medicine* aims to be at the forefront of this movement for sharing interdisciplinary knowledge and innovation in perioperative medicine on a global scale. With our talented editorial board and editorial team, we provide our authors with the best experience from submission to post publication. We look forward to receiving inspiring and exciting submissions to this journal as we, in symbiosis, shape the field of perioperative medicine.

Acknowledgments

I would like to acknowledge Shruti Kochhar, Managing Editor at *JMIR Perioperative Medicine*, for her valuable input in the drafting and editing of this article, and Rajitha Sivakumaran of JMIR Publications, for her contributions as the copyeditor of this article.

Conflicts of Interest

NR is the Editor-in-Chief of *JMIR Perioperative Medicine*.

References

1. Jackson C, Morton W. Improvement in surgical operations. US patent 4848. 1846 Nov 12. URL: https://patents.google.com/patent/US4848A/en [accessed 2023-11-14]
2. Macewen W. General observations on the introduction of tracheal tubes by the mouth, instead of performing tracheotomy or laryngotomy. Br Med J 1880 Jul 24;2(1021):122-124 [FREE Full text] [doi: 10.1136/bmj.2.1021.122] [Medline: 20749630]
3. Engström C. Respirator designed according to a new principle. Sven Lakartidn 1953 Mar 06;50(10):545-552. [Medline: 13077245]
4. Lee AJ. The anaesthetic out-patient clinic. Anaesthesia 1949 Oct;4(4):169-174. [doi: 10.1111/j.1365-2044.1949.tb05837.x] [Medline: 18140320]
5. Delisle M, Pradarelli JC, Panda N, Koritsanszky L, Sonnay Y, Lipsitz S, et al. Variation in global uptake of the Surgical Safety Checklist. Br J Surg 2020 Jan;107(2):e151-e160. [doi: 10.1002/bjs.11321] [Medline: 31903586]
6. COVIDSurg Collaborative. Mortality and pulmonary complications in patients undergoing surgery with perioperative SARS-CoV-2 infection: an international cohort study. Lancet 2020 Jul 04;396(10243):27-38 [FREE Full text] [doi: 10.1016/S0140-6736(20)31182-X] [Medline: 32479829]

XSL•FO
RenderX

7. Milne-Ives M, Leyden J, Maramba I, Chatterjee A, Meinert E. The potential impacts of a digital preoperative assessment service on appointments, travel-related carbon dioxide emissions, and user experience: case study. JMIR Perioper Med 2022 Feb 16;5(1):e28612 [FREE Full text] [doi: 10.2196/28612] [Medline: 35171104]
8. Cole JH, Highland KB, Hughey SB, O'Shea BJ, Hauert T, Goldman AH, et al. The association between borderline dysnatremia and perioperative morbidity and mortality: retrospective cohort study of the American College of Surgeons National Surgical Quality Improvement program database. JMIR Perioper Med 2023 Mar 16;6:e38462 [FREE Full text] [doi: 10.2196/38462] [Medline: 36928105]
9. Sjoding MW, Dickson RP, Iwashyna TJ, Gay SE, Valley TS. Racial bias in pulse oximetry measurement. N Engl J Med 2020 Dec 17;383(25):2477-2478 [FREE Full text] [doi: 10.1056/NEJMc2029240] [Medline: 33326721]
10. Ye J. Patient safety of perioperative medication through the lens of digital health and artificial intelligence. JMIR Perioper Med 2023 May 31;6:e34453 [FREE Full text] [doi: 10.2196/34453] [Medline: 37256663]
11. Kendale S, Bishara A, Burns M, Solomon S, Corriere M, Mathis M. Machine learning for the prediction of procedural case durations developed using a large multicenter database: algorithm development and validation study. JMIR AI 2023 Sep 8;2:e44909 [FREE Full text] [doi: 10.2196/44909]
12. Chen P, Chen L, Lin Y, Li G, Lai F, Lu C, et al. Predicting postoperative mortality with deep neural networks and natural language processing: model development and validation. JMIR Med Inform 2022 May 10;10(5):e38241 [FREE Full text] [doi: 10.2196/38241] [Medline: 35536634]
13. Ramaswamy P, Shah A, Kothari R, Schloemerkemper N, Methangkool E, Aleck A, et al. An accessible clinical decision support system to curtail anesthetic greenhouse gases in a large health network: implementation study. JMIR Perioper Med 2022 Dec 08;5(1):e40831 [FREE Full text] [doi: 10.2196/40831] [Medline: 36480254]
14. Nepogodiev D, Martin J, Biccard B, Makupe A, Bhangu A, National Institute for Health Research Global Health Research Unit on Global Surgery. Global burden of postoperative death. Lancet 2019 Feb 02;393(10170):401. [doi: 10.1016/S0140-6736(18)33139-8] [Medline: 30722955]

Abbreviations

mHealth: mobile health

Edited by T Leung; submitted 06.11.23; this is a non–peer-reviewed article;accepted 07.11.23; published 21.11.23.

Please cite as:
Rohatgi N
JMIR Perioperative Medicine: A Global Journal for Publishing Interdisciplinary Innovations, Research, and Perspectives
JMIR Perioper Med 2023;6:e54344
URL: https://periop.jmir.org/2023/1/e54344
doi:10.2196/54344
PMID:37988142

©Nidhi Rohatgi. Originally published in JMIR Perioperative Medicine (http://periop.jmir.org), 21.11.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Perioperative Medicine, is properly cited. The complete bibliographic information, a link to the original publication on http://periop.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Teaching Basic Surgical Skills Using a More Frugal, Near-Peer, and Environmentally Sustainable Way: Mixed Methods Study

Ben Smith[1*], BSc (Hons), MBChB; Christopher Paton[1*], BSc (Hons), MBBS; Prashanth Ramaraj[1,2*], BSc (Hons), MBBS, PGCert, MSc, MRCS

[1]Department of General Surgery and Urology, Frimley Health Foundation Trust, Frimley, Surrey, United Kingdom

[2]Imperial College School of Medicine, Imperial College London, London, United Kingdom

[*]all authors contributed equally

Corresponding Author:
Prashanth Ramaraj, BSc (Hons), MBBS, PGCert, MSc, MRCS
Imperial College School of Medicine
Imperial College London
Imperial College Road
South Kensington
London, SW7 2AZ
United Kingdom
Phone: 44 020 7589 5111
Email: prashanth.ramaraj@nhs.net

Abstract

Background: The Royal College of Surgeons Basic Surgical Skills (BSS) course is ubiquitous among UK surgical trainees but is geographically limited and costly. The COVID-19 pandemic has reduced training quality. Surveys illustrate reduced logbook completion and increased trainee attrition. Local, peer-led teaching has been shown to be effective at increasing confidence in surgical skills in a cost-effective manner. Qualitative data on trainee well-being, recruitment, and retention are lacking.

Objective: This study aims to evaluate the impact of a novel program of weekly, lunchtime BSS sessions on both quantitative and qualitative factors.

Methods: A weekly, lunchtime BSS course was designed to achieve the outcomes of the Royal College of Surgeons BSS course over a 16-week period overlapping with 1 foundation doctor rotation. All health care workers at the study center were eligible to participate. The study was advertised via the weekly, trust-wide information email. Course sessions included knot tying, suturing, abscess incision and drainage, fracture fixation with application of plaster of Paris, joint aspirations and reductions, abdominal wall closure, and basic laparoscopic skills. The hospital canteen sourced unwanted pig skin from the local butcher for suturing sessions and pork belly for abscess and abdominal wall closure sessions. Out-of-date surgical equipment was used. This concurrent, nested, mixed methods study involved descriptive analysis of perceived improvement scores in each surgical skill before and after each session, over 4 iterations of the course (May 2021 to August 2022). After the sessions, students completed a voluntary web-based feedback form scoring presession and postsession confidence levels on a 5-point Likert scale. Qualitative thematic analysis of voluntary semistructured student interview transcripts was also performed to understand the impact of a free-to-attend, local, weekly, near-peer teaching course on perceived well-being, quality of training, and interest in a surgical career. Students consented to the use of feedback and interview data for this study. Ethics approval was requested but deemed not necessary by the study center's ethics committee.

Results: There were 64 responses. Confidence was significantly improved from 47% to 73% (95% CI 15%-27%; $P<.001$; $t_{13}=5.3117$) across all surgical skills over 4 iterations. Among the 7 semistructured interviews, 100% (7/7) of the participants reported improved perceived well-being, value added to training, and positivity toward near-peer teaching and 71% (5/7) preferred local weekly teaching. Interest in a surgical career was unchanged.

Conclusions: This course was feasible around clinical workloads, resourced locally at next to no cost, environmentally sustainable, and free to attend. The course offered junior doctors not only a weekly opportunity to learn but also to teach. Peer-led, decentralized surgical education increases confidence and has a positive effect on perceptions about well-being and training. We hope to disseminate this course, leading to reproduction in other centers, refinement, and wide implementation.

XSL•FO
RenderX

(JMIR Perioper Med 2023;6:e50212) doi:10.2196/50212

KEYWORDS

basic surgical skills; frugal; low cost; peer-assisted learning; near-peer learning; environmentally sustainable; free; surgical education

Introduction

Background

The Basic Surgical Skills (BSS) course is offered by all 4 UK Royal Colleges of Surgeons (RCSs). BSS is highly encouraged and is considered as compulsory at the registrar or residency stage but can be completed at the senior medical student and intern level. Surgical skill outcomes form part of both the postgraduate surgical and medical student curricula [1,2]. The achievement of these skills at UK medical schools is poor, with surgical teaching weighted toward gowning, gloving, and consultant-led theoretical teaching [3].

Although these courses are "not treated as profit making vehicles," their high cost is underpinned by location and resource and instructor availability [4]. However, tutor qualification, regardless of previous experience, has been found to confer no difference in surgical skill performance [5]. Furthermore, 2 recent meta-analyses have demonstrated that near-peer learning (NPL) or peer-assisted learning (PAL) are very effective in teaching clinical and practical skills [6,7]. A UK surgical trainee typically spends £9105 (US $11,106.55) on courses. In the current financial climate, local courses can act as free or affordable ways to learn and consolidate skills [8].

There is evidence suggesting that local, intermittent, NPL or PAL surgical skill courses are effective, are affordable, and have high quality. A recent randomized controlled trial of 20 sessions (45 min each) led to a significant increase in surgical skill performance. This was significantly increased when taught by a near-peer tutor for knot tying, suturing, and simulated laparotomy [9]. Confidence in more advanced skills, such as arterial ligation, was demonstrated in a 4-session program using both PAL and faculty-led teaching [10]. A regional course led by senior medical students and junior doctors at a UK tertiary center also led to significant improvements in confidence across 8 skill domains [11].

Surgical education has suffered owing to the COVID-19 pandemic, with the cancellation of centralized courses, local training opportunities, and elective theater lists. The negative effect on the Annual Review of Competency Progression outcomes, training extension, and logbook completion is well documented [12]. A recent survey suggests that only 4 postgraduate surgical training posts in the United Kingdom meet the minimum quality standards [13], and the latest survey of UK surgical trainees illustrates a lack of training opportunities and the need for alternative modes of teaching [14]. Average yearly attrition from UK surgical training from 2016 to 2021 is 2.68% and has been increasing [15]. Furthermore, as affordability gains primacy as a health care metric, local, high-quality teaching represents an important surgical example of frugal innovation [16,17].

There is evidence suggesting that local PAL or NPL alternatives can be effective in teaching BSS. Although quantitative data exist, limited qualitative data exist regarding the effect of these courses on factors such as well-being, value added to training, or opinions about teaching modalities.

The objective of this study was to evaluate the impact of a novel 16-week program of weekly, lunchtime BSS sessions run at a UK district general hospital on both quantitative and qualitative factors.

Aims

The aims of this study included the following:

1. Evaluating the effect of a novel, 10-session program of weekly BSS sessions on student confidence in each taught skill before and after the session
2. Evaluating student opinion through semistructured interview in the following domains: *opinion about the course, impact on specialty choice, value added to surgical placement, effect on well-being, opinion about PAL* and *preference of weekly versus intensive short teaching program*

Methods

Course Structure and Delivery

A 10-session curriculum based on the BSS course (Multimedia Appendix 1) was taught over a 16-week period, corresponding with a foundation year rotation for 4 iterations from May 2021 to August 2022. Sessions were delivered weekly for 2 hours either by a consultant, near peer, or peer tutor over lunchtime. The course structure is available in Multimedia Appendix 1.

The hospital's canteen was contacted, and the chef kindly sourced pig skin and pork belly not suitable for sale, free of charge, as tissue models for this course. This was procured alongside the standard biweekly hospital meat delivery. A room was booked in advance at the hospital and free of charge. When unavailable, the doctors' mess was used. Out-of-date surgical equipment was sourced from the operating theaters. Sessions on abscess drainage or cyst excision and laparoscopic skills were taught over 2 weeks.

The course was publicized to all employees of the hospital via the weekly, trust-wide emails. The decision was made to include all clinical and nonclinical staff of any grade and from any specialty, with each session being "walk-in" in style. A lead instructor was confirmed in advance from the general surgery, urology, vascular, and orthopedics junior and consultant surgeon email directory, and a lead facilitator was assigned per session from volunteer foundation doctors rotating through surgical departments.

The role of the lead instructor was to introduce the session, ensure that learning objectives were made clear and achieved, ensure that session timings were met, and assist near-peer

XSL•FO
RenderX

teachers. The role of the facilitator was to ensure that session materials were obtained from the hospital canteen, a room was booked, and web-based feedback forms had been created.

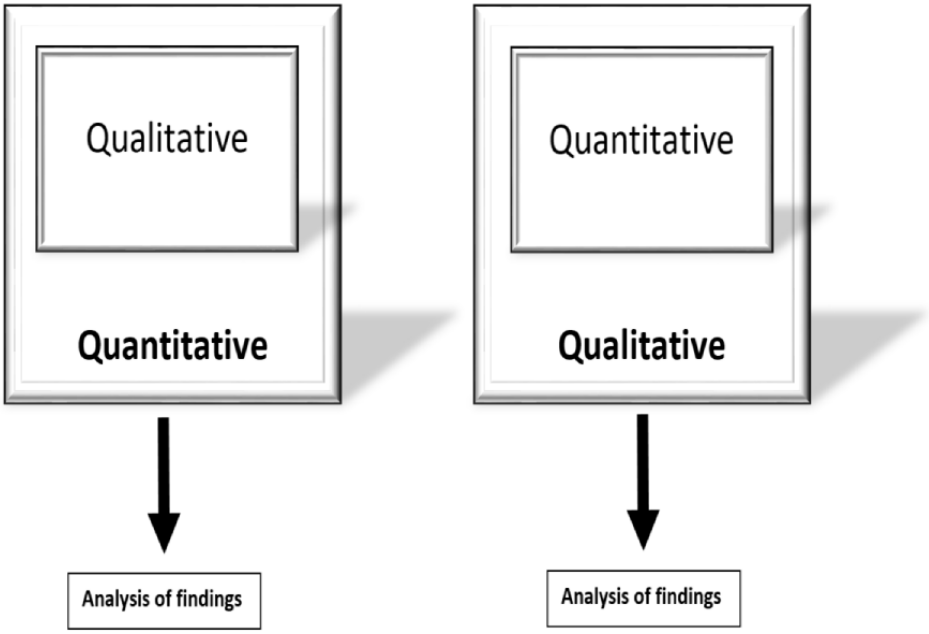
Feedback was sought on a voluntary basis in a paper-free manner from all participants before and after the sessions. An example feedback form is available in Multimedia Appendix 2. Google Forms were used in conjunction with QR codes. A Google integration was used to automate the population and distribution of attendance certificates to students upon completion of the feedback form as an incentive.

Study Design

A mixed methods approach using a concurrent nested design as per the typology by Plano Clark and Creswell [18] and mixed methods reporting guideline recommendations was used (Figure 1) [19]. This was used to provide a broad perspective than simply using an approach and allow qualitative investigation of effects that were not previously investigated such as perceived well-being.

The purpose was to gauge the impact of the course on the workforce's perceived confidence in surgical skills, resilience, and motivation toward a career in surgery.

Figure 1. Illustration of a concurrent, nested, mixed methods design.



Quantitative Analysis

A Google Form (Multimedia Appendix 2) was used to collect feedback from students after each session and is outlined in the following section:

1. Grade: health care assistant, nurse, operating department practitioner, medical student, foundation year or equivalent, core trainee or equivalent, specialty trainee or equivalent, or other
2. Confidence
   - Before the session: strongly disagree, disagree, neutral, agree, or strongly agree (supervised and unsupervised)
   - After the session: strongly disagree, disagree, neutral, agree, or strongly agree (supervised and unsupervised)
   - For those who answered either "strongly disagree" or "disagree," there was an additional question—"How many more sessions of 'X' do you feel you would need to attend to become confident to perform X skill independently?"
3. Session factors
   - Well structured: strongly disagree, disagree, neutral, agree, or strongly agree
   - Adequate supervision: strongly disagree, disagree, neutral, agree, or strongly agree
   - Materials were adequate for my learning: strongly disagree, disagree, neutral, agree, or strongly agree

Data are presented as a raw numbers and percentages. Percentages are rounded to the nearest integer.

Presession and postsession confidence analysis was conducted using the SPSS Statistics (version 28.0.1; IBM Corp) package. Percentage overall confidence was defined as the sum of the number of "confident" and "very confident" responses for "under supervision" and "without supervision" divided by the sum of the number of all responses for both "under supervision" and "without supervision." Overall confidence across the course was calculated as the sum of overall confidence per session divided by the number of sessions. The percentage of students who stated that they were "confident" or "very confident" to perform the skill before each session was compared with the percentage of students who were "confident" or "very confident" to perform the skill after each session using the paired 1-tailed *t* test. The Shapiro-Wilk test was used to test percentage confidence for normalcy. The null hypothesis (normal distribution) was accepted for overall percentage confidence (α=.05; *P*=.24; W=0.95 where α is the probability of type 1 error), confidence to perform the skill under supervision (α=.05; *P*=.07; W=0.88), and confidence to perform the skill independently (α=.05; *P*=.98; W=0.98). The difference between

XSL•FO
RenderX

the percentage of students who were "confident" or "very confident" to perform the skill before versus after each session was measured to calculate the difference across the whole course. This was performed for student confidence in performing the skills under supervision, independently, and overall. Data from laparoscopic skill sessions were not compared owing to heterogeneity in data collection among course iterations.

Qualitative Analysis

Students were retrospectively contacted via email and asked to complete a semistructured interview covering six areas, with an open question and Likert-scale or preference-based question for each (Multimedia Appendix 3):

1. General opinion about the BSS course
2. Impact on choice of specialty and why
3. Value added to junior doctor rotation and in what way
4. Effect on well-being
5. Opinion about near-peer teaching versus faculty-led teaching
6. Preference of weekly versus consolidated program for BSS

Thematic analysis was performed by authors, BS and CP, by coding major themes and subthemes from free-text answers and quantifying the number of responses related to each theme, following guidance from literature [20]. Agreement was achieved by inclusion of themes if there was consensus between the 2 coding authors. If there was disagreement, the discrepancy was discussed among all 3 authors to achieve consensus on included themes. These data are described as raw numbers and percentages, as was the discrete data from Likert scales and preference questions.

Ethical Considerations

The research and ethics committee of the study center was contacted to request for ethics approval. The authors were informed that ethics approval was not required as this study does not involve patient data, and all study participants provided written consent for nonidentifiable feedback data to be used for the purposes of this study. Attendance certificates were provided to students as an incentive upon the completion of the feedback form.

Results

Quantitative Analysis

Student Composition

There were 64 responses over 4 rotations, mainly from foundation doctors (n=52, 81%), followed by medical students (n=6, 9%), core surgical trainees or equivalent trust grades (n=5, 8%), and a nurse (n=1, 2%).

Student Confidence

Confidence results are illustrated in Table 1. A full narrative description per session is included in Multimedia Appendix 4. Confidence improved after the sessions in every surgical skill taught, for both supervised and unsupervised confidence. The largest number of responses were in the knot tying, suturing, and joint aspiration sessions. The lowest responses were in the fracture reduction, plastering, and abdominal closure sessions.

For the joint aspiration sessions, feedback was provided for individual joint (wrist and knee) in the second rotation compared with overall confidence for the session as a whole for rotations 1 and 3. There were 6 responses for confidence in wrist aspiration and 6 responses for confidence in knee aspiration, giving 12 data entries for this session. Combined with 2 responses in other rotations, in which feedback was not categorized based on joint, there were 14 total data entries.

For the laparoscopic skill session, data collection varied per rotation. Rotations 1 and 3 were measured based on skill, whereas rotation 2 was measured based on overall confidence—this is demonstrated in Table 1. The outcome for number of sessions required for confidence was omitted owing to inadequate completion.

Paired 1-tailed *t* test was used to measure the difference in confidence before and after the sessions. Across all sessions, except laparoscopic skills, which were excluded from analysis, the mean percentage of student responses in which students stated confidence to perform each skill supervised or unsupervised was 47%, whereas after the sessions, 73% of the students stated that they were confident to perform the skill supervised or unsupervised. The difference in confidence was 26% (95% CI 15%-27%; $P<.001$; $t_{13}=5.3117$).

The mean percentage of students who stated that they were confident in each skill increased the most in confidence to perform the skill under supervision with 60% before the sessions and 92% after the sessions (95% CI 12%-51%; $P=.008$; $t_{13}=3.9335$). The increase in confidence to perform each skill unsupervised was 20%; however, this was not statistically significant (34% before the sessions to 53% after the sessions; 95% CI 5%-44%; $P=.11$; $t_{13}=1.7435$).

XSL•FO
RenderX

Table 1. Student confidence before and after the teaching sessions.

Sessions and categories	Response before the sessions, n (%)					Response after the sessions, n (%)				
	Strongly disagree	Disagree	Neutral	Agree	Strongly agree	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
1: Knot tying (n=10)										
Supervised	0 (0)	3 (30)	1 (10)	2 (20)	4 (40)	0 (0)	0 (0)	0 (0)	5 (50)	5 (50)
Unsupervised	2 (20)	2 (20)	1 (10)	3 (30)	2 (20)	0 (0)	1 (10)	2 (20)	5 (50)	2 (20)
2: Suturing (n=15)										
Supervised	0 (0)	3 (20)	2 (13)	7 (47)	3 (20)	0 (0)	0 (0)	0 (0)	6 (40)	9 (60)
Unsupervised	3 (20)	2 (13)	2 (13)	5 (33)	3 (20)	0 (0)	0 (0)	6 (40)	4 (27)	5 (33)
3: Abscess drainage and cyst excision (n=7)										
Supervised	0 (0)	1 (14)	0 (0)	5 (71)	1 (14)	0 (0)	0 (0)	0 (0)	3 (43)	4 (57)
Unsupervised	0 (0)	2 (29)	3 (43)	1 (14)	1 (14)	0 (0)	2 (29)	2 (29)	1 (14)	2 (29)
4: Abdominal closure (n=5)										
Supervised	1 (20)	0 (0)	2 (40)	1 (20)	1 (20)	1 (20)	1 (20)	0 (0)	0 (0)	3 (60)
Unsupervised	0 (0)	1 (20)	2 (40)	2 (40)	0 (0)	0 (0)	0 (0)	1 (20)	4 (80)	0 (0)
5: Joint aspiration (n=14)										
Supervised	0 (0)	1 (7)	2 (14)	8 (57)	3 (21)	0 (0)	0 (0)	0 (0)	5 (36)	9 (64)
Unsupervised	1 (7)	6 (43)	5 (36)	1 (7)	1 (7)	0 (0)	1 (7)	9 (64)	3 (21)	1 (7)
6: Fracture reduction (n=4)										
Supervised	0 (0)	1 (25)	2 (50)	0 (0)	1 (25)	0 (0)	0 (0)	0 (0)	1 (25)	3 (75)
Unsupervised	1 (25)	2 (50)	1 (25)	0 (0)	0 (0)	0 (0)	0 (0)	3 (75)	1 (25)	0 (0)
7: Plastering (n=6)										
Supervised	1 (17)	1 (17)	0 (0)	2 (33)	2 (33)	1 (17)	0 (0)	0 (0)	2 (33)	3 (50)
Unsupervised	2 (33)	1 (17)	0 (0)	3 (50)	0 (0)	1 (17)	0 (0)	1 (17)	4 (67)	0 (0)
8a: Laparoscopic skills (by skill; n=3)										
Graspers: move objects between graspers	0 (0)	1 (33)	0 (0)	2 (67)	0 (0)	0 (0)	0 (0)	0 (0)	2 (67)	1 (33)
Graspers: stack dice	0 (0)	1 (33)	0 (0)	1 (33)	1 (33)	0 (0)	0 (0)	0 (0)	2 (67)	1 (33)
Scissors: cut shapes	0 (0)	1 (33)	1 (33)	1 (33)	0 (0)	0 (0)	0 (0)	0 (0)	3 (100)	0 (0)
Reef knot	2 (67)	0 (0)	0 (0)	1 (33)	0 (0)	0 (0)	0 (0)	2 (67)	1 (33)	0 (0)
8b: Laparoscopic skills (procedural confidence; n=3)										
Confidence in assisting laparoscopic procedures	1 (33)	0 (0)	0 (0)	2 (67)	0 (0)	0 (0)	0 (0)	1 (33)	1 (33)	1 (33)

XSL•FO
RenderX

Session Factors

The results for feedback questions related to session factors are illustrated in Table 2.

Each factor demonstrated mostly positive response, with only session structure obtaining a single negative response.

Table 2. Participant feedback results for session factors (structure, supervision, and teaching materials).

Factor	Response, n (%)				
	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Structure: the session was well structured (n=62)	1 (2)	0 (0)	4 (6)	30 (48)	28 (45)
Supervision: I felt adequately supervised (n=62)	0 (0)	0 (0)	5 (8)	21 (34)	36 (58)
Teaching materials: the teaching materials were adequate for my learning (n=59)	0 (0)	0 (0)	4 (7)	25 (42)	30 (51)

Qualitative Analysis

Semistructured Interviews

In total, 7 semistructured interviews were completed over 3 rotations. Refer to Multimedia Appendix 5 for figures outlining the full thematic analysis for each question.

Baseline Questions

Question 1: Were You Interested in Surgery as a Potential Career?

A slight majority of 57% (4/7) of attendees selected "yes," with 43% (3/7) selecting "no" for this question.

Question 2: Any Thoughts on Leaving Medicine Post Qualification?

Most attendees (5/7, 71%) at baseline answered "yes." Only 29% (2/7) had not had any thoughts about leaving medicine after qualification.

Question 3: What Was Your Overall Opinion of Your Surgical Rotation?

The overall opinion about surgical rotations among those surveyed was negative, with 71% (5/7) choosing "negative" and 29% (2/7) selecting "positive."

Main Questions

Main questions are as follows:

- Question 1—How did you find the basic surgical skills weekly course?
- Question 2—Did attending sessions have an impact on your choice of future specialty, if so in what way?
- Question 3—Did you find the sessions added value to your surgical placement/training, if so in what way?
- Question 4—Did attending practical sessions away from the ward influence your wellbeing and if so, how?
- Question 5—What's your opinion on peer assisted learning?
- Question 6—If you could be taught these skills during either a 16-week placement with weekly rostered sessions or in an intensive two days, which would you prefer and why?

Each question was accompanied by a Likert-scale response with the options "very negative/negative/neutral/positive/very positive" for questions 1 and 5; yes or no for question 2; "none, very little, moderately, significantly, greatly" for questions 3 and 4; and "16-weeks, neutral, 2 days" for question 6.

Thematic Codes and Excerpts

For question 1, a total of 5 major themes were identified: structure (1/7, 14%), usefulness (5/7, 71%), setting (2/7, 29%), availability (3/7, 43%), and content (3/7, 43%). The subthemes within "usefulness" were "practice surgical skills" (3/7, 43%) and "receiving training" (2/7, 29%). In the "availability" theme, "do not have to compete for theatre" (2/7, 29%) was most common, followed by "BSS course booked" (1/7, 14%). Regarding "content," the main subthemes were "covers BSS" (2/7, 29%) and "felt like a trainee" (1/7, 14%). The setting theme consisted of the "outside of theatre" (1/7, 14%) subtheme, and structure was "well-structured" (1/7, 14%).

The tone of all the free-text responses was positive. Following is an example excerpt:

> *Good – able to learn basic skills without fighting others for theatre time.*

There was only 1 free-text response to question 2, which was as follows:

> *More likely to choose surgery.*

Overall, 3 major themes were identified for question 3: "improving surgical skills" (3/7, 43%), "personal awareness" (1/7, 14%), and "improved training" (6/7, 86%). In the "improving surgical skills" theme, 3 subthemes were identified, such as "suturing" (1/7, 14%), "confidence to go to theatre" (1/7, 14%), and "practice outside of theatre" (1/7, 14%). Regarding "personal awareness," the specific context was "areas to improve" (1/7, 14%). In the "improved training" theme, the main subtheme was "only training received" (4/7, 57%), followed by "felt like being trained" (1/7, 14%) and "limited theatre time offered" (1/7, 14%).

Following is an excerpt:

> *Only departmental teaching – otherwise would be 100% service provision.*

Question 4 identified 3 major themes: "setting" (5/7, 71%), "distraction" (1/7, 14%), and "sensation of learning/improving" (6/7, 86%). Regarding "setting," the subthemes were "away from wards" (4/7, 57%) and "change on environment" (1/7, 14%). Regarding "distraction," the context was "attention to specific skills" (1/7, 14%). Regarding "sensation of learning/improving," the subthemes were "no other training" (4/7, 57%), "break from service provision" (1/7, 14%), and "acted as a break" (1/7, 14%).

XSL•FO
RenderX

Following is an excerpt:

> *Yes, by being away from the ward and paying attention to your training on specific skills you feel like you are improving and learning something.*

In total, 2 major themes were identified for question 5: "enjoyable" (7/7, 100%) and "pace" (1/7, 14%). In the "enjoyable" theme, the main subtheme was "learning from others" (5/7, 71%), followed by "teaching others" (2/7, 29%). Regarding "pace," the context was "able to go at own pace" (1/7, 14%).

Following is an excerpt:

> *Good to be able to learn from one another.*

For question 6, coding was split among those who preferred 2 days, those who preferred 16 weeks, and those who were neutral. Among those who preferred 2 days, the main themes were "easier" (2/7, 29%) and "density of learning" (1/7, 14%). Among those who preferred 16 weeks, responses were varied: "time for improvement" (2/7, 29%), "unable to attend all" (2/7, 29%), "able to develop questions" (1/7, 14%), "selectivity" (2/7, 29%), "enjoy weekly teaching" (2/7, 29%), "consolidate skills" (1/7, 14%), and "would not attend paid course" (1/7, 14%). The placement-dependent nature of the 14% (1/7) neutral responses is outlined in the following excerpt.

An attendee highlighted how the nature of the placement could influence this preference:

> *Depends on the placement. If your job is going to be very theatre based then it is better to do as two days at the start. If your job is more ward based then over 16 weeks provides time to practice skills and develop better questions to ask.*
>
> *Nice to be able to consolidate BSS knowledge from RCS course.*

Analysis of the Likert-Scale Responses to Each Main Question

Attendees had a positive experience overall, with 86% (6/7) rating their experience as "very positive" and 14% (1/7) rating it as "positive." Most attendees found that the course did not influence their choice of future specialty, with 86% (6/7) selecting "no." Only 14% (1/7) of the participants selected "yes." Overall, attendees found that the sessions added value, with 57% (4/7) selecting "greatly," 29% (2/7) selecting "significantly," and 14% (1/7) selecting "moderately." Overall, attendees found that the sessions improved well-being, with 57% (4/7) selecting "significantly" and 43% (3/7) selecting "greatly." Overall, attendees had a positive opinion about PAL, with 86% (6/7) rating their opinion as "very positive" and 14% (1/7) selecting "positive." Most attendees preferred a 16-week lunchtime course, with 71% (5/7) selecting this option and 14% (1/7) selecting "neutral" and "2 days," respectively.

Discussion

Principal Findings

This novel, 10-session BSS program improved confidence across all the surgical skills taught and improved student well-being through a series of structured practical sessions. The increase in student-reported overall confidence in skills taught across the course increased by 26% (95% CI 15%-27%; $P<.001$; $t_{13}=5.3117$). Most preferred to be taught over a 16-week period (4/7, 57%), enjoyed peer learning or NPL (7/7, 100%), and felt that it added value to the training (6/7, 86%).

Limitations

A limitation is that, for the joint aspiration sessions, the feedback was provided for individual joint (wrist and knee) in the second rotation compared with overall confidence for rotations 1 and 3, providing additional weighting to the second session in the results. Ideally, this would have been standardized and permitted the appreciation of joint-specific differences. Similarly, feedback for the laparoscopic skills session was incongruous owing to changes to the question posed on the feedback form between iterations. This would ideally have been standardized.

In future iterations of the course, organizers will endeavor to offer voluntary Objective Structured Assessments of Technical Skills to allow for the validation of the skills learned. The Objective Structured Assessments of Technical Skills scores could then be used to compare with those of the RCS BSS course to ensure the same standard.

In feedback-related data, there may be a degree of sampling bias as only outliers may wish to provide feedback. This was mitigated by automating the certificates of attendance once a feedback form was completed; therefore, all the students would complete the feedback forms to attain a certificate of attendance rather than just those who strongly wished to provide feedback. Self-selection bias is a feature in this cohort as only surgical workforce who were able to attend the sessions were able to provide feedback. Response bias was mitigated by anonymization of the feedback.

Comparison With Previous Studies

A significant barrier to facilitating teaching was senior clinician availability leading to junior clinician teaching. The results demonstrate that students valued PAL and were still able to gain confidence in surgical skills, consistent with the results of studies in a medical student cohort [10]. We believe this shows that senior clinician availability should not prevent students from teaching and illustrates how peer learning or NPL can lead to better teaching outcomes, particularly for practical skills [7].

Some strengths of this course were its frugality, use of local resource, and absence of cost for the participant. Most meat products used for BSS practice were waste offcuts from hospital kitchens. In addition, we used expired sutures and scalpels from the theater department. Our feedback forms were also digitalized using QR codes, and this meant that, once generated, only minimal changes needed to be made to the forms between rotations, and no paper waste was produced. This allowed the course to be organized by junior clinicians as it required little time during the day once the automation was set up. We believe this course is a strong example of frugal, environmentally sustainable innovation within surgical education. The average cost to run each course was £5 (US $6.13) to £10 (US $12.26) for pork belly for the abdominal wall closure session. Students

XSL•FO
RenderX

and educators were not required to travel across the country or further to attend a much-in-demand RCS BSS course.

The weekly nature of the course received very positive feedback, with students noting how it gave them "time for improvement" and did not penalize them if they were unable to attend all the sessions, without having to arrange study leave. Before BSS, students found the course as a useful primer, and after BSS, students found that this provided time for consolidation and an opportunity to teach colleagues.

There is good evidence to suggest that surgical skill teaching in medical school does not instill graduates with confidence in their practical skills such as suturing and knot tying [3]. This cohort has been the focus of previous studies such as that by Down et al [10], who demonstrated in a medical student cohort that BSS teaching by senior students (near peers) could improve confidence and suggested that this model could be taken forward into foundation training. When compared with our course, all 10 sessions improved student confidence in supervised procedures, consistent with previous studies evaluating the use of near-peer tutors for BSS teaching [9,10], and we have now shown how this is also applicable to a large cohort of recent graduates. In agreement with Pinter et al [9], this course also found high attendance and response rates from our knot tying and suturing sessions (10-15 students) compared with other sessions (≤7 students), which may suggest that students found the most value when practicing the frequently used BSS.

Furthermore, our use of near-peer tutors appears to initially support the work of Kim et al [5], demonstrating that the use of skill tutors can lead to similar learning outcomes compared with qualified surgeons when BSS are taught.

One of the major barriers to the RCS BSS course is the financial cost in the region of £650 (US $792.66), not inclusive of travel or accommodation costs. RCS BSS local centers price the course independently thus 1 center may charge £550 (US $670.67), whereas another may charge £850 (US $1,036.49). This novel course was free, local, and open to all clinical and nonclinical staff or students. Coupled with the increasing personal cost of surgical training, currently estimated to be >£9000 (>US $11,070) per junior doctor [8], these cost savings can make a significant difference to graduates and remove financial barriers to training.

This course will be refined and disseminated with data collected to demonstrate its effect in a large cohort. The aim is to expand this free-to-attend, low-cost, and local program to regional hospitals and collaborate nationally to try to improve the teaching of BSS. A list of recommendations for implementing and improving this course can be found at Multimedia Appendix 6.

Conclusions

In a training landscape of expensive, infrequent courses that often require significant travel and time out of hospital, our standardized, local BSS program can provide an effective, low-cost alternative with tangible benefits regarding both training value and student well-being.

Acknowledgments

The authors would like to thank all those who supported this novel program including instructors, students, administrative staff, and chefs. The authors would like to thank the Imperial College London Library for access to article processing charges (APCs) waiver through the Imperial-JMIR unlimited Open Access (OA) deal. This study was unfunded and received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data Availability

All data generated or analyzed during this study are included in this published paper (and the multimedia appendices). The authors will endeavor to provide any further relevant data upon reasonable request via email to author BS.

Authors' Contributions

BS, CP, and PR organized and executed the described course. PR led the conceptualization and methodology. BS, CP, and PR contributed to data curation equally. BS led the writing of the original draft manuscript, whereas CP and PR provided internal review. PR led the manuscript review and editing process. The corresponding author (PR) confirms on behalf of the authors that all those entitled to authorship have been included and no one who may be entitled has been excluded from authorship. All authors contributed equally and share first authorship.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Structure of the course.
[DOCX File , 12 KB - periop_v6i1e50212_app1.docx]

Multimedia Appendix 2

XSL•FO
RenderX

Example of the feedback form.
[DOCX File , 123 KB - periop_v6i1e50212_app2.docx]

Multimedia Appendix 3
Semistructured interview.
[DOCX File , 26 KB - periop_v6i1e50212_app3.docx]

Multimedia Appendix 4
Narrative summary of the quantitative results.
[DOCX File , 17 KB - periop_v6i1e50212_app4.docx]

Multimedia Appendix 5
Thematic analysis of the qualitative data.
[DOCX File , 30 KB - periop_v6i1e50212_app5.docx]

Multimedia Appendix 6
List of recommendations.
[DOCX File , 15 KB - periop_v6i1e50212_app6.docx]

References

1. Brecknell J, Lintott P. Curriculum core surgical training the intercollegiate surgical curriculum programme. Intercollegiate Surgical Curriculum Programme. 2021 Aug 04. URL: https://www.iscp.ac.uk/media/1371/core-surgical-training-curriculum-august-2021-version-2-july-2023.pdf [accessed 2021-11-09]
2. Outcomes for graduates. General Medical Council. 2020 Feb. URL: https://www.gmc-uk.org/education/standards-guidance-and-curricula/standards-and-outcomes/outcomes-for-graduates/outcomes-for-graduates [accessed 2021-11-09]
3. Davis CR, Toll EC, Bates AS, Cole MD, Smith FC. Surgical and procedural skills training at medical school - a national review. Int J Surg 2014 Aug;12(8):877-882 [FREE Full text] [doi: 10.1016/j.ijsu.2014.05.069] [Medline: 24909137]
4. Waduud MA, Ahmed N. Basic surgical skills course - why is it so costly? Ann Med Surg (Lond) 2015 Mar 20;4(2):98-99 [FREE Full text] [doi: 10.1016/j.amsu.2015.03.002] [Medline: 25870762]
5. Kim MJ, Boehler ML, Ketchum JK, Bueno R, Williams RG, Dunnington GL. Skills coaches as part of the educational team: a randomized controlled trial of teaching of a basic surgical skill in the laboratory setting. Am J Surg 2010 Jan;199(1):94-98. [doi: 10.1016/j.amjsurg.2009.08.016] [Medline: 20103072]
6. Guraya SY, Abdalla ME. Determining the effectiveness of peer-assisted learning in medical education: a systematic review and meta-analysis. J Taibah Univ Med Sci 2020 May 30;15(3):177-184 [FREE Full text] [doi: 10.1016/j.jtumed.2020.05.002] [Medline: 32647511]
7. Brierley C, Ellis L, Reid ER. Peer-assisted learning in medical education: a systematic review and meta-analysis. Med Educ 2022 May 24;56(4):365-373. [doi: 10.1111/medu.14672] [Medline: 34595769]
8. O'Callaghan J, Mohan HM, Sharrock A, Gokani V, Fitzgerald JE, Williams AP, Council of the Association of Surgeons in Training. Cross-sectional study of the financial cost of training to the surgical trainee in the UK and Ireland. BMJ Open 2017 Nov 15;7(11):e018086 [FREE Full text] [doi: 10.1136/bmjopen-2017-018086] [Medline: 29146646]
9. Pintér Z, Kardos D, Varga P, Kopjár E, Kovács A, Than P, et al. Effectivity of near-peer teaching in training of basic surgical skills - a randomized controlled trial. BMC Med Educ 2021 Mar 12;21(1):156 [FREE Full text] [doi: 10.1186/s12909-021-02590-2] [Medline: 33711985]
10. Down B, Morris S, Kulkarni S, Mohiuddin K. Effectiveness of a multi-session combined near-peer and faculty-led surgical skills course on self-perceived ability to perform basic surgical skills. Ann Med Surg (Lond) 2020 Jul 28;57:153-156 [FREE Full text] [doi: 10.1016/j.amsu.2020.07.045] [Medline: 32760585]
11. Bennett SR, Morris SR, Mirza S. Medical students teaching medical students surgical skills: the benefits of peer-assisted learning. J Surg Educ 2018 Nov;75(6):1471-1474. [doi: 10.1016/j.jsurg.2018.03.011] [Medline: 29653841]
12. Clements JM, Burke JR, Hope C, Nally DM, Doleman B, Giwa L, et al. The quantitative impact of COVID-19 on surgical training in the United Kingdom. BJS Open 2021 May 07;5(3):zrab051 [FREE Full text] [doi: 10.1093/bjsopen/zrab051] [Medline: 34169311]
13. Writing group, Project steering group, ASiT/BOTA Lost Tribe Study Group. Early years postgraduate surgical training programmes in the UK are failing to meet national quality standards: an analysis from the ASiT/BOTA Lost Tribe prospective cohort study of 2,569 surgical trainees. Int J Surg 2018 Apr;52:376-382 [FREE Full text] [doi: 10.1016/j.ijsu.2017.09.074] [Medline: 29037970]
14. JCST Trainee Survey. Joint Committee on Surgical Training. URL: https://www.jcst.org/quality-assurance/trainee-survey/ [accessed 2022-10-09]

XSL•FO
RenderX

15. Khalil K, Sooriyamoorthy T, Ellis R. Retention of surgical trainees in England. Surgeon 2023 Aug 21;21(4):203-207. [doi: 10.1016/j.surge.2022.09.001] [Medline: 36163150]
16. Bhatti Y, Taylor A, Harris M, Wadge H, Escobar E, Prime M, et al. Global lessons in frugal innovation to improve health care delivery in the United States. Health Aff (Millwood) 2017 Dec;36(11):1912-1919 [FREE Full text] [doi: 10.1377/hlthaff.2017.0480] [Medline: 29137503]
17. 17 PM, Bhatti Y, Harris M. Frugal and reverse innovations in surgery. In: Park A, Price R, editors. Global Surgery: The Essentials. Berlin, Germany: Springer Nature; 2017:193-206.
18. Plano Clark VL, Creswell JW. The Mixed Methods Reader. Thousand Oaks, CA: Sage Publications; 2008.
19. Lee SD, Iott B, Banaszak-Holl J, Shih SF, Raj M, Johnson KE, et al. Application of mixed methods in health services management research: a systematic review. Med Care Res Rev 2022 Jul 12;79(3):331-344. [doi: 10.1177/10775587211030393] [Medline: 34253078]
20. Castleberry A, Nolen A. Thematic analysis of qualitative research data: is it as easy as it sounds? Curr Pharm Teach Learn 2018 Jul;10(6):807-815. [doi: 10.1016/j.cptl.2018.03.019] [Medline: 30025784]

Abbreviations

BSS: Basic Surgical Skill
NPL: near-peer learning
PAL: peer-assisted learning
RCS: Royal College of Surgeon

Edited by T de Azevedo Cardoso, N Rohatgi; submitted 22.06.23; peer-reviewed by C Pyke, H Alsalhi; comments to author 05.09.23; revised version received 16.09.23; accepted 23.10.23; published 15.11.23.

Please cite as:
Smith B, Paton C, Ramaraj P
Teaching Basic Surgical Skills Using a More Frugal, Near-Peer, and Environmentally Sustainable Way: Mixed Methods Study
JMIR Perioper Med 2023;6:e50212
URL: https://periop.jmir.org/2023/1/e50212
doi:10.2196/50212
PMID:37966886

©Ben Smith, Christopher Paton, Prashanth Ramaraj. Originally published in JMIR Perioperative Medicine (http://periop.jmir.org), 15.11.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Perioperative Medicine, is properly cited. The complete bibliographic information, a link to the original publication on http://periop.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Survey of the Impact of Decision Support in Preoperative Management of Anemia (i-Anemia): Survey Study

Gaëtan Mignanelli[1*], MD; Richard Boyer[2*], MD, PhD; Nicolas Bonifas[3*], PhD; Emmanuel Rineau[4*], MD, PhD; Yassine Moussali[5*], MD; Morgan Le Guen[6*], MD, PhD

[1]Department of Anesthesia and Pain Medicine, Hôpital Foch, Université Versailles Saint Quentin, Neuilly sur Seine, France
[2]Department of Anesthesiology, Weill Cornell Medical College, New York, NY, United States
[3]Ecole Polytechnique, Palaiseau, France
[4]Department of Anesthesia and Critical Care, Angers University Hospital, University of Angers, Angers, France
[5]Department of Anaesthesia, Claude Galien Clinic, Quincy-sous-Sénart, France
[6]Department of Anesthesia and Pain Medicine, Hôpital Foch, Université Versailles Saint Quentin, Suresnes, France
[*]all authors contributed equally

Corresponding Author:
Gaëtan Mignanelli, MD
Department of Anesthesia and Pain Medicine
Hôpital Foch
Université Versailles Saint Quentin
51 Rue Edouard Nortier
Neuilly sur Seine, 92200
France
Phone: 33 0646288339
Email: gaetanms@gmail.com

Abstract

Background: Major surgery on patients with anemia has demonstrated an increased risk of perioperative blood transfusions and postoperative morbidity and mortality. Recent studies have shown that integrating preoperative anemia treatment as a component of perioperative blood management may reduce blood product utilization and improve outcomes in both cardiac and noncardiac surgery. However, outpatient management of anemia falls outside of daily practice for most anesthesiologists and is probably weakly understood.

Objective: We conducted a simulated case survey with anesthesiologists to accomplish the following aims: (1) evaluate the baseline knowledge of the preoperative optimization of anemia and (2) determine the impact of real-time clinical decision support on anemia management.

Methods: We sent a digital survey (i-Anemia) to members of the French Society of Anaesthesia and Critical Care. The i-Anemia survey contained 7 simulated case vignettes, each describing a patient's brief clinical history and containing up to 3 multiple-choice questions related to preoperative anemia management (12 questions in total). The cases concerned potential situations of preoperative anemia and were created and validated with a committee of patient blood management experts. Correct answers were determined by the current guidelines or by expert consensus. Eligible participants were randomly assigned to control or decision support groups. In the decision support group, the primary outcome measured was the correct response rate.

Results: Overall, 1123 participants were enrolled and randomly divided into control (n=568) and decision support (n=555) groups. Among them, 763 participants fully responded to the survey. We obtained a complete response rate of 65.6% (n=364) in the group receiving cognitive aid and 70.2% (n=399) in the group without assistance. The mean duration of response was 10.2 (SD 6.8) minutes versus 7.8 (SD 5) minutes for the decision support and control groups, respectively (P<.001). The score significantly improved with cognitive aid (mean 10.3 out of 12, SD 2.1) in comparison to standard care (mean 6.2 out of 12, SD 2.1; P<.001).

Conclusions: Management strategies to optimize preoperative anemia are not fully known and applied by anesthesiologists in daily practice despite their clinical importance. However, adding a decision support tool can significantly improve patient care by reminding practitioners of current recommendations.

XSL•FO
RenderX

(JMIR Perioper Med 2023;6:e49186) doi:10.2196/49186

KEYWORDS

anemia; transfusion; patient blood management; preoperative optimization; preoperative; blood; decision support; randomized; case; survey; anesthesiologists; anesthesiologist; anesthesia; anesthesiology; professional development; digital health; surgery; perioperative

Introduction

Several studies have demonstrated the harmful aspects of blood transfusion during hospitalization. In the perioperative period and in a critical care setting, it was associated with risks and additional costs without certain benefit in the absence of active bleeding [1,2]. These analyses led to a restriction in red blood cell (RBC) prescriptions and a parallel effort to control bleeding and to determine restrictive hemoglobin thresholds for RBC transfusion. Current guidelines suggest a trigger of 7-8 g/dL, with studies consistently demonstrating noninferior or superior outcomes compared to more liberal approaches (eg, hemoglobin triggers of 9-10 g/dL) [3-7].

However, anemia has been identified as a modifiable risk factor for poor perioperative outcomes [8,9]. Preoperative anemia is relatively common, affecting 25% to 75% of patients with an increasing prevalence in older patients and those with cancer [10,11]. Consequently, the relationship between anemia and perioperative transfusion provides a related risk for perioperative morbidity and mortality [12-15]. Management of preoperative anemia has become a goal for the anesthesiologist before surgery, with some recommendations endorsing anemia treatment in patients undergoing surgery [16,17]. Patient blood management (PBM), as defined by the Society for the Advancement of Blood Management, is the timely application of evidence-based medical and surgical concepts designed to maintain hemoglobin concentrations, optimize hemostasis, and minimize blood loss in an effort to improve patient outcomes [18]. There are 3 categories of actions that describe PBM in practice: the optimization of RBC mass, the reduction of blood loss and bleeding, and the optimization of the patient's physiological tolerance toward anemia [19]. Patient-centered decision-making is crucial when determining an individualized management plan and involves communication of the risks and benefits. Assessment of iron status, storage, and synthetic capacity should be performed to offer the best therapeutic strategy, which requires common biological tests measuring factors such as serum iron level, ferritin level, transferrin saturation [20,21]. A preanesthesia visit gives the opportunity to check for the presence or absence of anemia or iron deficiency because it allows enough time to correct or improve this functional defect. However, the current literature is still unclear regarding various situations that anesthesiologists must commonly manage. Therefore, implementation and access to consensus are practical aspects of PBM. Development of digital tools is crucial for detection and helping the physician with making a decision. In the present situation, we can imagine a digital "decision aid" that combines different guidelines and possesses an actualized decision tree to optimize PBM. We wish to measure this knowledge through a national and digital audit to further develop such a tool through a digital device.

The objectives of this work were 2-fold. First, we aimed to evaluate if the strategies of preoperative and postoperative anemia optimization by anesthesiologists fitted with the current recommendations. Second, we aimed to determine the usefulness of decision support to face these frequent and diverse situations involving anemia.

Methods

Ethical Considerations

This study received Institutional Review Board approval (CERAR IRB00010254-2020-131 - June 16, 2020). All eligible participants submitted a digital form of informed consent.

Participants

Voluntary clinicians were recruited in 3 ways to participate to this web-based questionnaire study. First, 2000 flyers with a QR code were distributed during the 3 days of the national congress of the French Society of Anaesthesia and Intensive Care (SFAR) in September 2021. Usually, this congress includes about 5000 to 6000 participants, including physicians, residents, and anesthetic nurses. Second, in the 2 months following the congress, we sent an email via the SFAR network to all members with a brief explanation and the same link. In parallel, the SFAR and the Association of Young Anesthesiologists (AJAR) shared the link on their social networks (Facebook, Instagram, and Twitter [subsequently rebranded as X]). All clinicians recruited were invited to share the link to the questionnaire with other anesthesiologists.

Protocol and Web Questionnaire

Once on the website, care providers selected whether or not to participate in this study. The first web page explained the aim of the study and reminded the reader of inclusion and exclusion criteria. Participants included only French-speaking anesthesiologists, including senior on-going physicians, fellows, and residents. Assistants, specialized nurse anesthesiologists, and medical students were not allowed to participate in the study.

If the participant was eligible, a digital form of informed consent was submitted. There was a brief explanation of the study, which specified that only anesthesiologists could participate and that the results would be anonymized. They agreed that their data would be anonymously used. We also provided the reference number of the Institutional Review Board approval so the participants could check it if they wanted. Then, access to the questionnaire was opened via a public web portal using an internet-based polling software called Typeform [22,23], which allowed participants to anonymously complete the questionnaire from different devices or navigators. Only 1 access per participant was available based on the participant's IP address. Once consent was obtained, the participants approved their

XSL•FO
RenderX

email to finally be randomized. Some questions about demographics were asked at this step.

The survey was created and validated by a committee of PBM experts. The final survey included 7 short clinical case vignettes containing up to 3 multiple-choice questions each (totaling 12 questions). These cases specifically involved potential preoperative and postoperative anemia situations. Some of them were built according to existing guidelines, while others led to an open discussion due to a lack of adapted recommendations. Moreover, no questions about intraoperative management of patients with anemia or control of bleeding were proposed to focus on the perioperative course (Table 1). Following receipt of informed consent, participants were randomly and equally divided into 2 groups: one with decision support and one without (the control group). The decision support group benefited from cognitive assistance during the survey; this was not available to the control group. The cognitive assistance consisted of the current recommendations about anemia management. This assistance was only visible if an error was made in the clinician's responses. The clinician then had the choice to select a different answer or not.

Table 1. Summary of the questionnaire's questions and objectives.

Case	Subject of the question	Specialty	Objective	References
1	Definition of preoperative hemoglobin thresholds	Orthopedics	Define the basic knowledge	[17]
1	Treatment of anemia (iron therapy and EPO[a])	Orthopedics	Define the management of anemia	[17]
1	Indication to change an oral iron therapy with bad tolerance by an intravenous therapy	Orthopedics	Management of side effects	[17]
2	Woman without anemia; no indication to begin any treatment	Gastrointestinal surgery	Define the basic knowledge	[24]
3	Poorly tolerated anemia; indication to transfuse 1-2 blood units	Gynecological surgery	Define the management of anemia	[25]
3	Indication to start IV[b] iron therapy to restore the iron status	Gynecological surgery	Define the management of anemia	[17]
4	Indication to postpone surgery	Gastrointestinal surgery	Define the management of anemia	[25]
5	Definition of preoperative hemoglobin thresholds for patients with terminal renal failure	Chronic kidney failure	Define the basic knowledge	[26]
5	Treatment of anemia (iron therapy first, then EPO if persisting anemia)	Chronic kidney failure	Define the management of anemia	[26]
5	Stop supplementations before polycythemia (thrombosis risk)	Chronic kidney failure	Management of side effects	[26]
6	EPO and iron therapy before surgery	Urgent cardiac surgery	Define the management of anemia	[27]
7	IV iron therapy for persisting iron deficiency after 34 weeks of amenorrhea	Obstetrics	Define the management of anemia	[28]

[a]EPO: erythropoietin.

[b]IV: intravenous.

Data Collected

The following demographic characteristics of the population were collected: age, experience, specialties, and site of practice (public, private, or mixed). The time to complete the survey was recorded. Moreover, the quality of the responses was analyzed (false or correct). Cognitive assistance consisted of the current recommendations about anemia management. In fact, if a participant in the decision support group gave a wrong answer, a window appeared showing the current recommendations and their references (Table 1). The participants then had the choice to follow the recommendations and change their answer or to continue with their original answer. In the control group, participants were not able to go back and change answers to previous questions during the study. At the end of the survey, scores were made available to the participants. Individual participant data will not be shared.

Statistical Analysis

The number of participants was calculated to allow for interpretation of the results using parametric tools. The main hypothesis was superiority in the rate of correct responses among physicians in the decision support group, with a planned difference of 25%. The estimated rate of correct response in the control group was 66%. With an α score of 5% and a power of 90%, at least 110 physicians were required. Moreover, to ensure a national representation of experience and geography, we determined the required number of completed forms to be 500 out of 9000 active physicians.

The data are expressed as the median (IQR) and mean (SD) according to their nonnormal or normal distribution. The quality

XSL•FO
RenderX

of the questionnaire was determined based on the percentage of correct responses, and the comparison was mainly performed using chi-square and Wilcoxon-Mann-Whitney tests, as appropriate. The analysis was performed using SPSS 12.0 (IBM).

Results

In total, 1939 people opened the link, of which 1123 participants were randomized into the decision support and control groups before completing the questionnaire. Specifically, 555 (49.4%) anesthesia care providers were place in the decision support group receiving cognitive aid, and 568 (50.6%) were placed in the control group. The final cohort included 763 active French-native anesthesiologists representing the different fields of care (private and public). These participants included senior fellows and residents, as expected (Figure 1 and Table 2), for 763 questionnaires totally completed.

Figure 1. Flowchart of the i-Anemia participants.

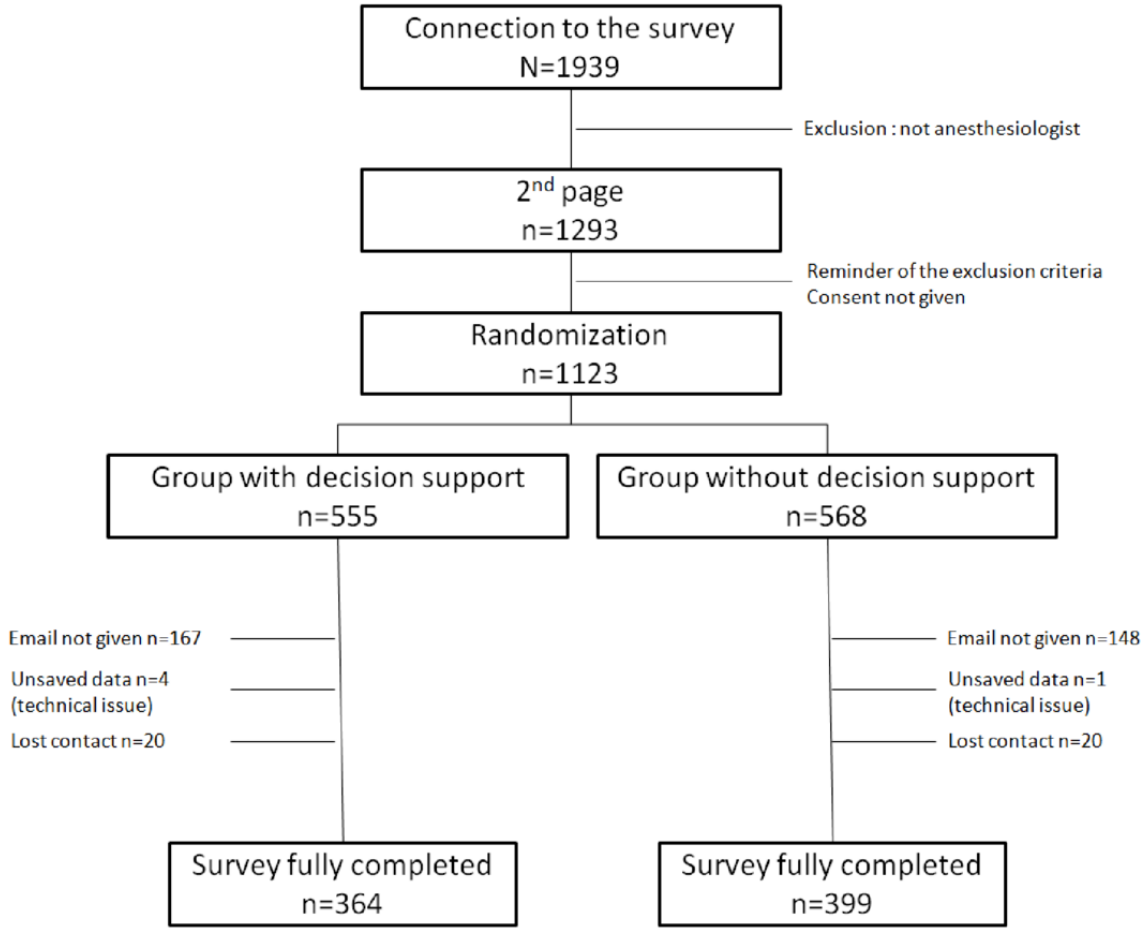


Table 2. Characteristics of the survey participants.

Characteristic, n (%)	All participants (n=1123)	With decision support (n=555)	Without decision support (n=568)
Type of health care center			
University tertiary hospital	347 (30.9)	181 (16.1)	166 (14.8)
Public secondary hospital	194 (17.3)	102 (9.1)	92 (8.2)
Miscellaneous (cancer centers, mixed organization)	109 (9.7)	60 (5.3)	49 (4.4)
Private care center	238 (21.2)	119 (10.6)	119 (10.6)
Anesthesia residents	232 (20.7)	106 (9.5)	126 (11.2)
Device used to participate			
Smartphone	804 (71.6)	380 (33.8)	424 (37.8)
Computer	317 (28.2)	152 (13.5)	165 (14.7)
Miscellaneous (eg, tablet)	2 (0.2)	2 (0.2)	0 (0)

XSL•FO
RenderX

The response rate was 65.6% (n=364) in the decision support group and 70.2% (n=399) in the control group (*P*=.49). Only respondents who fully completed the questionnaire were analyzed in the main outcome. The mean response times were 12.1 (SD 21.7) minutes and 9.2 (SD 23.5) minutes (*P*<.001) for the decision support and control groups, respectively. If we remove the outliers, defined as people who took more than 1 hour to complete the survey (n=5 in the decision support group and n=2 in the control group), the average time to answer was 10.2 (SD 6.8) minutes and 7.8 (SD 5.0) minutes (*P*<.001) for the decision support and control groups, respectively. The rate of correct responses was significantly higher in the decision support group (mean 10.3, SD 2.1) than in the control group (mean 6.2, SD 2.1; *P*<.001) (Table 3).

Table 3. Distribution of correct responses to each of the 12 multiple-choice questions for the decision support and control groups.

Question	Correct response rate, n (%)	
	Decision support (n=364)	Control (n=399)
Q1	321 (88.2)	221 (55.4)
Q2	295 (81)	178 (44.6)
Q3	309 (85)	241 (60.3)
Q4	320 (88.1)	306 (76.7)
Q5	350 (96.1)	363 (90.9)
Q6	350 (96.1)	288 (72.2)
Q7	298 (81.8)	146 (36.7)
Q8	331 (90.9)	133 (33.4)
Q9	307 (84.3)	20 (5.1)
Q10	312 (85.7)	235 (59)
Q11	318 (87.3)	128 (32.2)
Q12	319 (87.6)	150 (37.5)

Analyzing each multiple-choice question, the results were significantly better in the group with decision support (mean 87.7%, SD 4.6%) compared to the control group (mean 50.3%, SD 22.5%) (*P*<.001). Another interesting result was the relevant reduction in the variability of responses in the decision support group, which suggests that a guided response helps prevent participants from moving forward with a false response. This difference was more important in situations with unclear recommendations or in uncommon or specific cases (eg, questions 7-9 and 11-12). Question 7 was about postponing an elective surgery with mild risk of bleeding when anemia is discovered late (Multimedia Appendix 1). Questions 8 and 9 were about end stage renal disease (ESRD) and major surgery. Question 11 was about major vascular surgery, and question 12 asked about the use of intravenous iron therapy in obstetrics. These results show a great discrepancy between the two groups. The decision support group gave a more homogeneous response (*P*<.001), resulting in a better mean score compared to the control group (*P*<.001; Table 4). The magnitude of the difference was independent of the kind of institution and the experience level of anesthesiologist (senior or junior).

XSL•FO
RenderX

Table 4. Distribution of total scores out of 12 at the end of the survey for participants in the decision support and control groups.

Score	Participants, n (%)	
	Decision support (n=364)	Control (n=399)
0	2 (0.5)	0 (0)
1	1 (0.3)	3 (0.8)
2	0 (0)	11 (2.8)
3	0 (0)	24 (6)
4	2 (0.5)	43 (10.8)
5	7 (1.9)	66 (16.5)
6	4 (1.1)	68 (17)
7	19 (5.2)	80 (20.1)
8	19 (5.2)	55 (13.8)
9	31 (8.5)	32 (8)
10	57 (15.7)	11 (2.8)
11	80 (22)	6 (1.5)
12	142 (39)	1 (0.3)

Discussion

Principal Findings

This survey demonstrated that better performance in the management of preoperative anemia can be achieved if the anesthesiologists, regardless of experience, have at their disposal a cognitive aid ($P<.001$). Usually, guidelines require a certain time before being implemented, and decision support may accelerate this shift and help physicians make a decision, especially in rare situations or situations with a low level of prior knowledge.

The World Health Organization defines PBM as “a patient-focused, evidence-based, and systematic approach to optimize the management of patients and transfusion of blood products for quality and effective patient care. It is designed to improve patient outcomes through the safe and rational use of blood and blood products and by minimizing unnecessary exposure to blood products” [29]. Indeed, the incidence of preoperative anemia (about 30% regardless of the type of surgery) is critically important and increases with age. Additionally, preoperative anemia, which represents an independent risk factor of morbidity and mortality, may delay functional recuperation with a major risk of loss of autonomy in older patients [8,30]. The interest in the preoperative field is the potential reversibility of this factor if recognized with a different line of therapy, as included in different guidelines.

Despite their clinical and physiological fundamental importance, the strategies for optimizing the management of perioperative anemia are not always well-known and applied by anesthesiologists in daily practice. It is essential to understand the underlying reasons to improve practices and quality of care. This was the primary aim of the present study, and we observed weak knowledge of the recommendations among active anesthesiologists, with a mean rate of 6.2 (SD 2.1) out of 12 correct responses.

The second conclusion is the potential value of an immediately available cognitive aid, which significantly improved the rate of correct responses in our study. Several reasons may explain the moderate adherence to guidelines about perioperative anemia management. First, the importance of PBM is a relatively new topic that requires visibility and is not well-known by clinicians. In fact, the occurrence of complications in patients with anemia are mid-term or long-term complications, and the delay between general anesthesia and occurrence of complications may disconnect the physician from the patient’s daily analysis and prescriptions. Moreover, at the time of our study, some situations had been noted in the last recommendation, especially in the Frankfurt Consensus Conference published in 2018, and there was probably a need for updating to include additional situations, such as chronic renal failure and functional surgery [17,25]. Finally, strategies of PBM take time to be fully efficient, sometimes requiring a few weeks, and if the experts suggest postponing an elective surgery to correct iron deficiency or anemia, this is still difficult in many departments due to the typical course of patient care, the need to respect surgical planning, and the patient’s needs and wishes. To postpone an elective surgery remains difficult if the parties involved (eg, surgeon and anesthesiologist) do not share the same view of the importance of anemia management, and this is still the case in many teams. Furthermore, these strategies represent additional costs in the patients’ pathway, even if they reduce morbidity and mortality on balance [31].

The other aspect concerns the implementation of a cognitive aid. On one hand, there is communication work to be carried out to formulate clear and applicable recommendations in daily practice, which is a work in progress [32]. On the other hand, the addition of a decision support tool significantly improved patient management by reminding practitioners of the current recommendations. A simple line of work to help improve the quality of care could then be an implementation of these kinds of aids within the software used by anesthesiologists for their

XSL•FO
RenderX

consultations, for example. Interest in such a solution was recently explored regarding the decision to perform transfusion. One single center randomized controlled trial compared young physicians who received computerized decision support about transfusion to those who did not [33] and showed an increase in the appropriate transfusion rate (RBCs, platelets, fresh frozen plasma) in the computerized decision support group (40.4%) compared with the control group (32.5%; risk ratio 1.24, 95% CI 1.13-1.37). Additionally, 3 cohort studies assessed RBC usage before and after the intervention [34-36] and showed a significant reduction in overall or inappropriate RBC usage (RBC transfusions per 100 inpatient days, $P<.001$) after computerized decision support was implemented. A statistically significant reduction in overall or inappropriate RBC usage over time ($P=.01$) was also reported. Furthermore, reduced 30-day re-admission (5.2%; risk ratio 0.62, 95% CI 0.56-0.69) and mortality (2.2%; risk ratio 0.60, 95% CI 0.51-0.71) were found in one single center trial. This cognitive reminder could be applied not only to anesthesiologists but to surgeons or any specialists involved in the medical management of the patients to share the crucial role of optimization. In view of its demonstrable benefits, there is an increasing awareness of the need to integrate the pillars of PBM within routine surgical care. In the United States, PBM has been successfully introduced in some centers [37], while in western Australia, it has become the standard of care [38]. PBM initiatives in Europe, however, have been variable and inconsistent, reflecting the difficulties that can be met with its implementation [39]. Multiple barriers, including lack of knowledge, interdisciplinary commitment, resources, and concerns, limit the translation of PBM guidelines into clinical practice [33-35,40]. Education and training should be considered to increase awareness of the clinical implications of anemia and the need for alternatives to transfusion. Typically, the lowest rate of correct responses in the control group concerned the management of anemia in patients with chronic kidney failure experiencing some trouble with iron supplements, erythropoietin, and situations with hemoglobin levels greater than 11 g/dL and high ferritin. Other cases with a low rate of correct responses were orthopedic cases and the management of anemia for pregnant women. Certain situations are well-described in the guidelines, but some patients fall in grey areas, warranting the use of cognitive aids to make the most pertinent choice. Once any deficiencies in the underlying knowledge have been addressed, attitude and behavior should change [41]. All recommendations and standard operating procedures must be easily accessible and aimed at supporting clinical judgment as the cornerstone of patient care.

The main population who completed our survey was composed of 763 active anesthesiologists; this is a large sample size with a high rate of complete answers and represents all our care areas and experience levels. This was permitted by using different channels of diffusion. We created the cases with the help of experts in PBM and made them representative of real-life, everyday clinical practice, which is an important strength.

However, there are several limitations of our study. The questionnaire study was destined only for French participants. Furthermore, there was a bias in participant recruitment, and we can assume a worse rate of correct responses among nonresponding physicians or incomplete questionnaires. Third, we did not track the rate of response changes in the decision support group to objectively measure the effect of receiving help. Additionally, we could not verify that all the participants truly were anesthesiologists; if they declared that they were we approved their participation without any further control. Finally, recommendations evolve quickly, which may have misled some participants.

Conclusion

The recommendations about current PBM are not well-known by practicing physicians. We found that the implementation of decision support is very useful to face frequent and diverse situations of anemia and to improve the quality of patient management. Collaborative and continuous efforts to translate PBM guidelines into clinical practice, if done in an engaged, multidisciplinary, organized, and structured way, could make PBM the norm.

Acknowledgments

We thank Raphael Martin, Xavier Capdevila, and Jean-Marc Constantin of the French Society of Anaesthesia and Intensive Care and Charles Marc Samama for their contributions.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Case reports and questions.
[DOCX File , 15 KB - periop_v6i1e49186_app1.docx]

References

1. Shander A, Hofmann A, Ozawa S, Theusinger OM, Gombotz H, Spahn DR. Activity‐based costs of blood transfusions in surgical patients at four hospitals. Transfusion 2010 Mar 25;50(4):753-765. [doi: 10.1111/j.1537-2995.2009.02518.x]
2. Vincent JL, Baron JF, Reinhart K, Gattinoni L, Thijs L, Webb A, et al. Anemia and blood transfusion in critically ill patients. JAMA 2002 Sep 25;288(12):1499-1507. [doi: 10.1001/jama.288.12.1499] [Medline: 12243637]

XSL•FO
RenderX

3. Hébert PC, Wells G, Blajchman MA, Marshall J, Martin C, Pagliarello G, et al. A multicenter, randomized, controlled clinical trial of transfusion requirements in critical care. N Engl J Med 1999 Feb 11;340(6):409-417. [doi: 10.1056/nejm199902113400601]
4. Carson JL, Terrin ML, Noveck H, Sanders DW, Chaitman BR, Rhoads GG, et al. Liberal or restrictive transfusion in high-risk patients after hip surgery. N Engl J Med 2011 Dec 29;365(26):2453-2462. [doi: 10.1056/nejmoa1012452]
5. Villanueva C, Colomo A, Bosch A, Concepción M, Hernandez-Gea V, Aracil C, et al. Transfusion strategies for acute upper gastrointestinal bleeding. N Engl J Med 2013 Jan 03;368(1):11-21. [doi: 10.1056/nejmoa1211801]
6. Murphy GJ, Pike K, Rogers CA, Wordsworth S, Stokes EA, Angelini GD, et al. Liberal or restrictive transfusion after cardiac surgery. N Engl J Med 2015 Mar 12;372(11):997-1008. [doi: 10.1056/nejmoa1403612]
7. Hajjar L, Vincent JL, Galas FRBG, Nakamura R, Silva C, Santos MH, et al. Transfusion requirements after cardiac surgery: the TRACS randomized controlled trial. JAMA 2010 Oct 13;304(14):1559-1567. [doi: 10.1001/jama.2010.1446] [Medline: 20940381]
8. Fowler A, Ahmad T, Phull M, Allard S, Gillies M, Pearse R. Meta-analysis of the association between preoperative anaemia and mortality after surgery. Br J Surg 2015 Oct;102(11):1314-1324 [FREE Full text] [doi: 10.1002/bjs.9861] [Medline: 26349842]
9. Ranucci M, Di Dedda U, Castelvecchio S, Menicanti L, Frigiola A, Pelissero G. Impact of preoperative anemia on outcome in adult cardiac surgery: a propensity-matched analysis. Ann Thorac Surg 2012 Oct;94(4):1134-1141. [doi: 10.1016/j.athoracsur.2012.04.042] [Medline: 22698773]
10. Shander A, Knight K, Thurer R, Adamson J, Spence R. Prevalence and outcomes of anemia in surgery: a systematic review of the literature. Am J Med 2004 Apr 05;116 Suppl 7A:58S-69S. [doi: 10.1016/j.amjmed.2003.12.013] [Medline: 15050887]
11. Guralnik J, Eisenstaedt R, Ferrucci L, Klein H, Woodman R. Prevalence of anemia in persons 65 years and older in the United States: evidence for a high rate of unexplained anemia. Blood 2004 Oct 15;104(8):2263-2268 [FREE Full text] [doi: 10.1182/blood-2004-05-1812] [Medline: 15238427]
12. Karkouti K, Wijeysundera DN, Beattie WS, Reducing Bleeding in Cardiac Surgery (RBC) Investigators. Risk associated with preoperative anemia in cardiac surgery: a multicenter cohort study. Circulation 2008 Jan 29;117(4):478-484. [doi: 10.1161/CIRCULATIONAHA.107.718353] [Medline: 18172032]
13. Herzog CA, Muster HA, Li S, Collins AJ. Impact of congestive heart failure, chronic kidney disease, and anemia on survival in the Medicare population. J Card Fail 2004 Dec;10(6):467-472. [doi: 10.1016/j.cardfail.2004.03.003] [Medline: 15599836]
14. Rosencher N, Kerkkamp HE, Macheras G, Munuera L, Menichella G, Barton DM, OSTHEO Investigation. Orthopedic surgery transfusion hemoglobin European overview (OSTHEO) study: blood management in elective knee and hip arthroplasty in Europe. Transfusion 2003 Apr;43(4):459-469. [doi: 10.1046/j.1537-2995.2003.00348.x] [Medline: 12662278]
15. Lu M, Sing DC, Kuo AC, Hansen EN. Preoperative anemia independently predicts 30-Day complications after aseptic and septic revision total joint arthroplasty. J Arthroplasty 2017 Sep;32(9S):S197-S201. [doi: 10.1016/j.arth.2017.02.076] [Medline: 28390884]
16. Society of Thoracic Surgeons Blood Conservation Guideline Task Force, Ferraris VA, Brown JR, Despotis GJ, Hammon JW, Reece TB, et al. 2011 update to the Society of Thoracic Surgeons and the Society of Cardiovascular Anesthesiologists blood conservation clinical practice guidelines. Ann Thorac Surg 2011 Mar;91(3):944-982. [doi: 10.1016/j.athoracsur.2010.11.078] [Medline: 21353044]
17. Mueller MM, Van Remoortel H, Meybohm P, Aranko K, Aubron C, Burger R, et al. Patient blood management: recommendations from the 2018 Frankfurt Consensus Conference. JAMA 2019 Mar 12;321(10):983-997. [doi: 10.1001/jama.2019.0554] [Medline: 30860564]
18. Gani F, Cerullo M, Ejaz A, Gupta P, Demario V, Johnston F, et al. Implementation of a blood management program at a tertiary care hospital: effect on transfusion practices and clinical outcomes among patients undergoing surgery. Ann Surg 2019;269(6):1073-1079. [doi: 10.1097/sla.0000000000002585]
19. Shander A, Van Aken H, Colomina M, Gombotz H, Hofmann A, Krauspe R, et al. Patient blood management in Europe. Br J Anaesth 2012 Jul;109(1):55-68 [FREE Full text] [doi: 10.1093/bja/aes139] [Medline: 22628393]
20. Sweitzer BJ. Preoperative screening, evaluation, and optimization of the patient's medical status before outpatient surgery. Curr Opin Anaesthesiol 2008;21(6):711-718. [doi: 10.1097/aco.0b013e3283126cf3]
21. Goodnough L, Nemeth E, Ganz T. Detection, evaluation, and management of iron-restricted erythropoiesis. Blood 2010 Dec 02;116(23):4754-4761 [FREE Full text] [doi: 10.1182/blood-2010-05-286260] [Medline: 20826717]
22. Typeform. URL: https://www.typeform.com [accessed 2023-11-13]
23. Bienvenue dans l'étude clinique iAnemie. URL: https://hoopcare.typeform.com/to/euMcGDXF?typeform-source=ianesthesie.typeform.com [accessed 2023-11-13]
24. Haemoglobin concentrations for the diagnosis of anaemia and assessment of severity. World Health Organization. 2011. URL: https://www.who.int/publications/i/item/WHO-NMH-NHD-MNM-11.1 [accessed 2023-11-13]
25. Muñoz M, Acheson AG, Auerbach M, Besser M, Habler O, Kehlet H, et al. International consensus statement on the peri‐operative management of anaemia and iron deficiency. Anaesthesia 2016 Dec 20;72(2):233-247. [doi: 10.1111/anae.13773]

XSL•FO
RenderX

26. KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease. Kdigo. URL: https://kdigo.org/wp-content/uploads/2017/02/KDIGO_2012_CKD_GL.pdf [accessed 2023-11-13]
27. Task Force on Patient Blood Management for Adult Cardiac Surgery of the European Association for Cardio-Thoracic Surgery (EACTS)the European Association of Cardiothoracic Anaesthesiology (EACTA), Boer C, Meesters MI, Milojevic M, Benedetto U, Bolliger D, et al. 2017 EACTS/EACTA guidelines on patient blood management for adult cardiac surgery. J Cardiothorac Vasc Anesth 2018 Feb;32(1):88-120 [FREE Full text] [doi: 10.1053/j.jvca.2017.06.026] [Medline: 29029990]
28. Muñoz M, Peña-Rosas JP, Robinson S, Milman N, Holzgreve W, Breymann C, et al. Patient blood management in obstetrics: management of anaemia and haematinic deficiencies in pregnancy and in the post-partum period: NATA consensus statement. Transfus Med 2018 Feb;28(1):22-39 [FREE Full text] [doi: 10.1111/tme.12443] [Medline: 28722245]
29. WHO global forum for blood safety: patient blood management. World Health Organization. 2022. URL: https://www.who.int/news-room/events/detail/2011/03/14/default-calendar/who-global-forum-for-blood-safety-patient-blood-management [accessed 2023-11-13]
30. Musallam KM, Tamim HM, Richards T, Spahn DR, Rosendaal FR, Habbal A, et al. Preoperative anaemia and postoperative outcomes in non-cardiac surgery: a retrospective cohort study. The Lancet 2011 Oct;378(9800):1396-1407. [doi: 10.1016/s0140-6736(11)61381-0]
31. Kaserer A, Rössler J, Braun J, Farokhzad F, Pape H, Dutkowski P, et al. Impact of a patient blood management monitoring and feedback programme on allogeneic blood transfusions and related costs. Anaesthesia 2019 Dec;74(12):1534-1541 [FREE Full text] [doi: 10.1111/anae.14816] [Medline: 31448406]
32. Gestion du capital sanguin en pré, per et post opératoire et en obstétrique. Haute Autorité de Santé. URL: https://www.has-sante.fr/jcms/p_3193968/fr/gestion-du-capital-sanguin-en-pre-per-et-post-operatoire-et-en-obstetrique [accessed 2023-11-13]
33. Rothschild JM, McGurk S, Honour M, Lu L, McClendon AA, Srivastava P, et al. Assessment of education and computerized decision support interventions for improving transfusion practice. Transfusion 2007 Feb;47(2):228-239. [doi: 10.1111/j.1537-2995.2007.01093.x] [Medline: 17302768]
34. Adams E, Longhurst C, Pageler N, Widen E, Franzon D, Cornfield D. Computerized physician order entry with decision support decreases blood transfusions in children. Pediatrics 2011 May;127(5):e1112-e1119. [doi: 10.1542/peds.2010-3252] [Medline: 21502229]
35. Goodnough LT, Maggio P, Hadhazy E, Shieh L, Hernandez-Boussard T, Khari P, et al. Restrictive blood transfusion practices are associated with improved patient outcomes. Transfusion 2014 Oct;54(10 Pt 2):2753-2759. [doi: 10.1111/trf.12723] [Medline: 24995770]
36. Kassakian SZ, Yackel TR, Deloughery T, Dorr DA. Clinical decision support reduces overuse of red blood cell transfusions: interrupted time series analysis. Am J Med 2016 Jun;129(6):636.e13-636.e20. [doi: 10.1016/j.amjmed.2016.01.024] [Medline: 26873112]
37. Theusinger O, Kind S, Seifert B, Borgeat L, Gerber C, Spahn D. Patient blood management in orthopaedic surgery: a four-year follow-up of transfusion requirements and blood loss from 2008 to 2011 at the Balgrist University Hospital in Zurich, Switzerland. Blood Transfus 2014;12(2):203. [doi: 10.1007/978-3-642-55004-1_13]
38. Verdecchia NM, Wisniewski MK, Waters JH, Triulzi DJ, Alarcon LH, Yazer MH. Changes in blood product utilization in a seven-hospital system after the implementation of a patient blood management program: a 9-year follow-up. Hematology 2016 Sep;21(8):490-499. [doi: 10.1080/10245332.2015.1112496] [Medline: 26870887]
39. Xydas S, Magovern CJ, Slater JP, Brown JM, Bustami R, Parr GV, et al. Implementation of a comprehensive blood conservation program can reduce blood use in a community cardiac surgery program. J Thorac Cardiovasc Surg 2012 Apr;143(4):926-935 [FREE Full text] [doi: 10.1016/j.jtcvs.2012.01.003] [Medline: 22285327]
40. Yaffee DW, Smith DE, Ursomanno PA, Hill FT, Galloway AC, DeAnda A, et al. Management of blood transfusion in aortic valve surgery: impact of a blood conservation strategy. Ann Thorac Surg 2014 Jan;97(1):95-101. [doi: 10.1016/j.athoracsur.2013.09.057] [Medline: 24263014]
41. De Leeuw NK, Lowenstein L, Hsieh YS. Iron deficiency and hydremia in normal pregnancy. Medicine (Baltimore) 1966 Jul;45(4):291-315. [doi: 10.1097/00005792-196607000-00002] [Medline: 5946689]

Abbreviations

PBM: patient blood management
RBC: red blood cell
SFAR: French Society of Anaesthesia and Intensive Care

XSL•FO
RenderX

Edited by T de Azevedo Cardoso, T Aslanidis; submitted 20.05.23; peer-reviewed by A Myrou, S Hidig; comments to author 05.07.23; revised version received 23.08.23; accepted 05.11.23; published 01.12.23.

Please cite as:
Mignanelli G, Boyer R, Bonifas N, Rineau E, Moussali Y, Le Guen M
Survey of the Impact of Decision Support in Preoperative Management of Anemia (i-Anemia): Survey Study
JMIR Perioper Med 2023;6:e49186
URL: https://periop.jmir.org/2023/1/e49186
doi:10.2196/49186
PMID:

©Gaëtan Mignanelli, Richard Boyer, Nicolas Bonifas, Emmanuel Rineau, Yassine Moussali, Morgan Le Guen. Originally published in JMIR Perioperative Medicine (http://periop.jmir.org), 01.12.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Perioperative Medicine, is properly cited. The complete bibliographic information, a link to the original publication on http://periop.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org

XSL•FO
RenderX