# 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 9 (2026) ISSN 2561-9128 Editor in Chief: Nidhi Rohatgi, MS, MD, SFHM

## Contents

### Review

Multidimensional Assessment of Recovery After Total Knee Arthroplasty in Clinical Practice: Critical Narrative Review (e84011)
Abderrahmane Boukabache, Nimalan Maruthainar, Vikrant Manhas, Darren Player. . . . . 3

### Original Papers

Forced-Air Warming Temperature Settings for Treating Postoperative Hypothermia in the Postanesthesia Care Unit: Randomized Controlled Trial (e85045)
Koravee Pasutharnchat, Rattaphol Seangrung, Sirikarn Sirisophaphong, Wilailuck Wongkum. . . . . 19

Virtual Reality for the Management of Postoperative Pain and Anxiety in Children and Adolescents Undergoing Nuss Repair of Pectus Excavatum: Randomized Controlled Trial (e80902)
Charlotte Walter, Dillon Froass, Nora Bell, Lauren Haack, Chloe Boehmer, Claudia Bruguera Torres, Rachel Spivak, Max Chou, Kristie Geisler, Keith O'Conor, Sara Williams, Lili Ding, Christopher King, Vanessa Olbrecht. . . . . 32

Evaluating the Impact of Virtual Reality on Orthopedic Trauma Skills Acquisition Among Surgical Residents: Randomized Crossover Study (e79343)
Sandipan Chatterjee, Khairul Mohammad, Monica Ghidinelli. . . . . 46

A Novel Customizable Datamart and Tableau Dashboard to Monitor Multiple Enhanced Recovery After Surgery Programs: Development and Validation Study (e82472)
Sunitha Singh, Susannah Oster, Efrat Bolze, Aaron Sasson, James Nicholson, Elliott Bennett-Guerrero. . . . . 56

Comparison Between Ultrasound and Magnetic Resonance Imaging Measurements of the Optic Nerve Sheath Diameter in Patients Undergoing Intracranial Surgery: Prospective Observational Single-Center Study (e67480)
Mauricio Giraldo, Luz Lopera, Aly Balbaa, Nelson Gonzalez, Raffael Cezar Zamper, Michael Mayich, Mel Boulton. . . . . 69

Survival Prediction in Patients With Bladder Cancer Undergoing Radical Cystectomy Using a Machine Learning Algorithm: Retrospective Single-Center Study (e86666)
Francesco Causio, Vittorio De Vita, Andrea Nappi, Melissa Sawaya, Bernardo Rocco, Nazario Foschi, Giuseppe Maioriello, Pierluigi Russo. .
7 7

Physician Perspectives on ChatGPT-4o as a Patient Resource for Abdominal Cancer Surgeries: Cross-Sectional Survey (e81374)
Christina Lindsay, Devika Shenoy, Allison Martin, Christie Clipper, Kevin Shah, Michael Lidsky, Daniel Nussbaum, Ralph Snyderman. . . . . 96

AI-Generated Avatar Videos for Postoperative Patient Education Among Health Care Workers: Pilot Randomized Controlled Trial (e89277)
Syed Haider, Srinivasagam Prabha, Cesar Gomez-Cabello, Ariana Genovese, Bernardo Collaco, Nadia Wood, James London, Sanjay Bagaria, Mark Lifson, Cui Tao, Antonio Forte. . . . . 110

Enhancing the User Experience of a Perioperative Digital Health Tool for Information Exchange Using a Human-Centered Design Thinking Approach: Qualitative Observational Study (e79349)
Charlé Steyl, Carljohan Orre, Greg Foster, Hanel Duvenage, Michelle Chew, Hyla Kluyts. . . . . 126

Clinical Feasibility and Outcomes of Surgeon-Performed Laparoscopic-Guided Subcostal Transversus Abdominis Plane Block in Laparoscopic Cholecystectomy: Prospective Observational Study (e87622)
Sarun Mahasupachai, Thawatchai Tullavardhana. . . . . 147

Immersive Virtual Reality for Pain and Relaxation in Older Adults Following Elective Inpatient Abdominal Surgery: Single-Arm Study Examining Feasibility and Acceptability (e81791)
Christina Keny, Ujala Shafiq, Karl Lorenz, Marcia Russell, Heather Leutwyler, Laura Wagner, Victoria Tang, Linda Park. . . . . 156

Association Between Complications and Death Within 30 Days After Orthopedic Surgery: Vascular Events in Noncardiac Surgery Patients Cohort Evaluation (VISION) Substudy (e90823)
Lily Park, P Devereaux, Ameen Patel, Vikas Tandon, Diane Heels-Ansdell, Lehana Thabane, Pablo Serrano, Matthew Chan, Wojciech Szczeklik, Sadeesh Srinathan, Ignacio Garutti, Gerard Urrutia, Ernesto Guerra-Farfan, Hassaan Abdel Khalik, Emmanuelle Duceppe, Sandra Ofori, Maura Marcucci, David Conen, Michael Wang, Jessica Spence, Daniel Tushinski, Kamal Bali, Anthony Adili, Vickas Khanna, Ana Tonelli, Francesca Mulazzani, Wenjun Jiang, Olufemi Ayeni, Gerard Slobegean, Theodore Miclau, Mohit Bhandari, Flavia Borges. . . . . 171

Assessing the Effects of eHealth Literacy and the Area Deprivation Index on Barriers to Electronic Patient Portal Use for Orthopedic Surgery: Cross-Sectional Observational Study (e72035)
Audrey Litvak, Nicholas Lin, Kelly Hynes, Jason Strelzow, Jeffrey Stepan. . . . . 185

Benchmark Development for Fundamental Arthroscopic Skills Using a Simulation-Based Training Program: Observational Study (e82723)
Eric Davis, Brianna Caraet, Robert Pedowitz, Gregg Nicandri. . . . . 196

# Multidimensional Assessment of Recovery After Total Knee Arthroplasty in Clinical Practice: Critical Narrative Review

Abderrahmane Boukabache[1], BSc (Hons), MSc; Nimalan Maruthainar[2], BSc (Hons), MBBS; Vikrant Manhas[3], MBBS, MS; Darren Player[1], BSc (Hons), PhD

[1]Division of Surgery and Interventional Science, Faculty of Medical Sciences, University College London, Bloomsbury Campus: 2nd Floor, Charles Bell House, 43-45 Foley Street, London, United Kingdom

[2]Department of Trauma and Orthopaedics, Royal Free Hospital, London, United Kingdom

[3]Department of Orthopaedics, All India Institute of Medical Sciences, New Delhi, India

**Corresponding Author:**
Abderrahmane Boukabache, BSc (Hons), MSc
Division of Surgery and Interventional Science, Faculty of Medical Sciences, University College London, Bloomsbury Campus: 2nd Floor, Charles Bell House, 43-45 Foley Street, London, United Kingdom

## Abstract

**Background:** Total knee arthroplasty (TKA) is the primary treatment for advanced knee osteoarthritis. Despite its clinical success and favorable patient-reported outcome measures (PROMs), approximately 20% to 30% of patients continue to experience persistent functional limitations and muscle weakness. This highlights the need for a comprehensive evaluation of recovery parameters beyond pain and range of motion. Given the wide range of methods available for assessing TKA outcomes, clinicians often select tools based on personal preference and understanding, which may affect accuracy and consistency; for example, the Knee Injury and Osteoarthritis Outcome Score may overestimate function compared to gait analysis studies.

**Objective:** The aim of this study was to conduct a narrative review focusing on the use, strengths, and limitations of different outcome measures used in routine orthopedic practice to optimize post-TKA evaluation.

**Methods:** A literature search was conducted in February 2025 across 2 databases (PubMed and Web of Science). Eligible studies included original research articles, systematic reviews, and meta-analyses that focused on validated measures used to evaluate TKA. Case reports, conference abstracts, and studies focused exclusively on surgical techniques were excluded. Themes were identified across studies to structure the results according to types of assessments and clinical applicability.

**Results:** A total of 6831 studies were retrieved and screened in this review, with 4 themes emerging around muscle mass, strength, performance, and PROMs. The Oxford Knee Score is favored for its ease of use and minimal ceiling effects. Broader tools like the Knee Injury and Osteoarthritis Outcome Score and Western Ontario and McMaster Universities Osteoarthritis Index provide detailed insights but are less practical clinically. For muscle strength, the portable fixed dynamometer showed high reliability and comparability to isokinetic dynamometry. Dual-energy X-ray absorptiometry remains the gold standard for assessing muscle mass, while bioelectrical impedance analysis offers a practical alternative. The 5-Repetition Sit-to-Stand test effectively evaluates lower limb power and speed.

**Conclusions:** Clinicians should integrate both objective (muscle mass, strength, and performance) and subjective (PROMs) measures to improve TKA recovery assessment. This multidimensional approach has the potential to enhance the accuracy of patient evaluation and supports the development of tailored rehabilitation strategies that address individual deficits and optimize functional outcomes.

***(JMIR Perioper Med 2026;9:e84011)*** doi:10.2196/84011

**KEYWORDS**

arthroplasty; osteoarthritis; function; outcome; total knee replacement

## Introduction

Osteoarthritis is a leading cause of pain and disability worldwide, affecting 528 million people as of 2019 [1,2]. Among various treatment options for osteoarthritis, total knee arthroplasty (TKA) stands out for its substantial impact on alleviating chronic knee pain [3]. With advancements in prosthetic design and surgical techniques, TKA has demonstrated high survival rates and relatively low complication risks, making it one of the most successful orthopedic procedures. However, despite these clinical successes, postoperative functional outcomes can vary among patients [4], with approximately 20% to 30% experiencing persistent functional limitations and muscle weakness [5].

Several factors influence recovery following TKA, including surgeon-related aspects such as surgical volume and technique, as well as patient-related variables like preoperative physical conditioning and psychological status. Among these, preoperative quadriceps strength has been identified as a key predictor of postoperative function, directly impacting mobility and the ability to perform daily activities [6]. Nonetheless, frequently used tools to evaluate TKA in orthopedics often rely on measures such as pain assessment and range of motion (ROM) [7], which, while valuable, may not fully capture postoperative recovery or functional capacity.

Evidence suggests that these metrics correlate poorly with objective measures of physical function and do not adequately reflect muscle weakness or biomechanical impairments that persist post-TKA [8]. While patient-related outcome measures (PROMs) provide subjective insights into perceived recovery, they may overestimate functional improvements compared to more objective assessments, such as gait analysis and performance-based tests [9]. This discrepancy underscores the need for a more comprehensive evaluation framework that integrates multiple dimensions of recovery, including muscle mass, muscle strength, and physical performance.

This review aims to critically examine the use, strengths, and limitations of different outcome measures used in routine practice when assessing recovery post-TKA, emphasizing the importance of incorporating objective physical performance metrics alongside PROMs. We hypothesize that a multidimensional approach, combining assessments of muscle mass, strength, physical performance, and PROMs, will provide a more accurate and clinically meaningful evaluation of TKA recovery, ultimately guiding more effective rehabilitation strategies.

## Methods

A literature search of studies focusing on methods assessing TKA outcomes was performed using PubMed and Web of Science databases in February 2025. No limits were applied to publication dates. Inclusion criteria comprised original research, systematic reviews, and meta-analyses. After the removal of duplicates and cross-referencing of articles, we screened titles and abstracts for relevance. Studies were excluded if they were case reports, conference abstracts, or focused exclusively on surgical techniques. Only studies using standardized and validated assessment tools were included to ensure clinical relevance. The methodological quality of included studies was evaluated based on study design, sample size, and clarity of outcome measures.

One author conducted the initial screening of all retrieved abstracts against the inclusion criteria. Full texts of potentially relevant studies were then independently reviewed for eligibility and data extraction. Discrepancies between reviewers were resolved through discussion until consensus was achieved. Extracted data were thematically organized (muscle mass, muscle strength, physical performance, and PROMs), allowing patterns across assessment domains to be identified and synthesized into a structured narrative aligned with the review's objectives.

Given the narrative nature of this review, our selection was not strictly systematic; instead, we documented the total number of retrieved and screened articles (n=6831).

## Results

### Patient Reported Outcome Measures

There are numerous PROMs available to assess the outcomes of TKA, including: Knee Society Clinical Rating System (KSS), Western Ontario and McMaster Universities Osteoarthritis (WOMAC), Knee Injury and Osteoarthritis Outcome Score (KOOS), and the Oxford Knee Score (OKS) [10-13]. In clinical practice and research, it is essential to prioritize the use of sound PROMs over frequently used ones. The ultimate goal herein is to find a balance between standardized measures and specific contextually relevant PROMs in the field of TKA.

The KSS and OKS PROMs are shorter than the KOOS and WOMAC, with 10 and 12 items, respectively. The OKS uses a 4-week recall period, while the patient-reported component of the KSS is designed to reflect current knee status without a fixed recall window, which may reduce recall bias but limit comparability across time points. The OKS is solely derived from patient input, while the KSS comprises a PROM section completed by the patient and an informational-clinical section for the surgeon. Only the PROM section of the KSS is used to produce a psychometrically valid knee score. This separation sought to allow the functional PROM section to be independently assessed from the clinical section and confounding factors such as age or comorbidities. Notably, Martimbianco et al [14] reported considerable inter- and intraexaminer variability in the clinical-reported section of the KSS, which raises concerns about its reliability and consistency in clinical use. In contrast, the KOOS and WOMAC are larger instruments, with 42 and 33 items, respectively. They assess symptoms, stiffness, pain, and activities of daily living (ADL), but also include additional domains such as sports or recreation and quality of life. Like the OKS, these instruments are totally focused on the patient and their self-reported experiences. The WOMAC uses a 48-hour recall period, which minimizes recall bias and improves measurement precision; however, this narrow timeframe may fail to reflect symptom variability and functional fluctuation over longer recovery periods following TKA. In contrast, the KOOS uses a 1-week recall period, providing a broader representation of knee symptoms and functional limitations. Nevertheless, this longer recall window may increase susceptibility to recall bias and potentially reduce sensitivity to short-term clinical change.

All 4 PROMs assess pain and activity function, but they differ in how they capture pain information. The OKS and KSS both include a single item that captures the patient's general level of pain. However, the OKS asks further pain information in the context of general activities, such as work interference and walking. This is particularly important to assess as movement is associated with changes in pain, from subtle changes in muscle coordination to complete avoidance of joint function [15]. In contrast, the KOOS and WOMAC both include a section dedicated to assessing pain experienced during specific activities, with nine identical items. This encourages the patient

to recall and report on the pain experience during knee maneuvers such as knee twisting and straightening. The KOOS is an extension of the WOMAC, which justifies the similarities in the items. It is used in younger and/or more active patients who typically perform demanding twisting movements in their ADL. By assessing knee pain during movements like twisting and straightening, changes in functional outcomes over time can be tracked. This is especially relevant post-TKA, where pain during specific activities is directly linked to the functionality of the joint and the surrounding muscles [16].

The only resemblance when assessing function across the 4 PROMs is the patient's comfort level when handling stair ascent and descent. This likely reflects the universal importance of these activities in daily life and the significant impact that knee function has on an individual's mobility and well-being. However, each PROM also has unique items that assess different aspects of knee function and quality of life.

The OKS, KOOS, and WOMAC all assess the knee concerning various ADL, such as kneeling, transportation, domestic work, and bathroom activities. The OKS and KSS share only the item related to the distance the patient can walk. The KSS and items within the symptoms section of both KOOS and WOMAC ask about the patient's ability to fully bend and extend the knee, as well as objective measures of flexion contracture and extension lag. The use of similar items and domains across these scores can lead to greater consistency in assessing knee function and symptoms. Furthermore, consistency is important for tracking changes over time and comparing outcomes across different patients and studies. The KOOS is unique in that it assesses ADL related to sports and recreation and includes items that ask about quality of life. While similarities can provide consistency, differences between PROMs may capture important nuances specific to each condition.

The KSS evaluates 4 distinct domains: clinical, functional, satisfaction, and fulfillment of expectation. A notable feature is the inclusion of both high-demand tasks and 3 patient-selected priority activities from a predefined list, aiming to tailor the assessment to individual goals and enhance the relevance of functional evaluation. This patient-led component is designed to offer a more individualized perspective on recovery and may help inform more targeted postoperative rehabilitation. However, despite these strengths, the validity of the KSS has been questioned, particularly in the context of TKA outcome measurement. Several studies have noted limitations, including Ghanem et al [17], who highlighted concerns in revision TKA scenarios, and Vogel et al [18], who found the system less responsive than broader tools like the 36-Item Short Form Survey and WOMAC. One key criticism lies in the item-selection process: patients were not involved in the development of the tool, which may result in content that does not fully reflect patient priorities or lived experiences. Bach et al [19] further noted that the relatively limited item pool may constrain the scale's ability to capture the full spectrum of functional outcomes. This lack of patient input and limited item scope could contribute to a misalignment between what the tool measures and what matters most to patients, ultimately impacting its content validity and clinical use [17,18,20].

Ghanem et al [17] highlight that 2 very different patients could receive the same clinical score on the KSS. For instance, a patient with a stiff, pain-free, well-aligned knee and a patient with mild pain, excellent knee motion, and normal alignment may both receive a similar score, even though they have very different levels of function in their daily lives. To address these concerns, a revised Knee Society Scoring System (2011 Knee Society Score) was developed and validated as a more reliable measure for assessing outcomes in TKA procedures [21,22]. Albeit, a revised outcome measure, PROMs have inherent limitations in that they rely primarily on self-reported data, which may not always align perfectly with objective measures of physical function. This is where muscle function analysis could come into play. It offers an objective assessment of a patient's physical capabilities, including muscle strength, range of motion, and muscle activation patterns. By quantifying these aspects, health care providers can validate and complement the patient's self-reported symptoms and limitations with tangible data. This objective information not only supports the patient's reported experiences but also provides a more complete picture of their health status.

Despite its validity issues, the KSS is still popular among clinical researchers [23], perhaps for including alignment and ROM measurement. Proper coronal alignment of the components in TKA has been shown in the literature to be critical for implant survival rate. Additionally, knee ROM is a crucial indicator of successful TKA and is necessary for many ADLs [24]. While ROM is commonly measured and reported, the significance of isolated muscle function has been overlooked here. Capin et al [25] found quadriceps strength in particular to be a critical factor in TKA recovery and considered a rate-limiting step. This limited quantification of isolated muscle function is an area that warrants attention and improvement in the care of patients who underwent TKA.

The OKS was developed specifically for evaluating the outcomes of TKA. The simplicity and shortness of the questionnaire make it an attractive option for clinicians, and this may in part have contributed to its broad use in cohort studies and joint replacement registers [26,27]. This intentional oversimplification of the questionnaire highlights a lack of scope and sophistication to adequately capture the complex, interrelated issues experienced by many patients, such as joint stiffness, muscle weakness, and instability. To address these limitations, the use of a combination of assessment tools and methods has been suggested [28]. This could include performance-based functional tests, isolated muscle function tests, muscle mass measurement, and imaging studies.

Another consideration is that OKS has a high response rate when compared to other PROMs [10]. Analysis of the National Health Service PROMs dataset of 72,154 OKS concluded that OKS does not exhibit a ceiling or a floor effect at 6 months [29]. Marx et al [30] reported a ceiling effect of 22% at 12 months following surgery, but this could be attributed to patients achieving an optimal outcome rather than a limitation of the OKS. Nevertheless, this could indicate that the OKS may be more appropriate for short-term outcomes and might inadequately reflect the long-term burden. Although it has been

used in randomized controlled trials to assess knee arthroplasty long-term outcomes [31].

The WOMAC, developed in 1982, has undergone multiple revisions since and has been validated for TKA clinical trials [32-34], with ceiling effects at 6 months and 12 months for patients who underwent TKA [35]. The KOOS was an upgrade to the WOMAC to effectively capture the requirements of younger and more active individuals with knee injury or osteoarthritis. Several studies have shown that the KOOS is more sensitive and responsive than the WOMAC in younger or more active patients with knee injury or knee osteoarthritis [36]. Roos and Toksvig-Larsen [35] evaluated the outcome of 105 patients (mean age of 71) after TKA and found 74% of all Sport/Recreation items were considered to be "not applicable." The floor effects of approximately 48% likely reflect the high-demand activities, such as sports and recreation, which may be more relevant for younger, more active adults undergoing TKA.

A systematic review of the literature performed by Collins et al [37] found, as it was intended, the KOOS's ADL subscale has better content validity for older patients, and Sport/Recreation has better content validity for younger patients. Increasing ceiling effects from 6 to 12 months, particularly in the pain domain, may suggest that some patients have reached a plateau in their recovery, but also raises questions about the sensitivity of the measurement tool.

According to the Consensus-Based Standards for the Selection of Health Measurement Instruments (COSMIN), which define internationally accepted criteria for evaluating the validity, reliability, and responsiveness of PROMs, the OKS meets COSMIN requirements and is recommended for use as a TKA outcome measure. Pain and function subscales of WOMAC and KOOS also demonstrate adequate measurement properties when evaluated as standalone subscales. However, the KSS does not consistently fulfill all COSMIN criteria and is not classified among the instruments meeting minimum standards for psychometric validation [38]. Several studies have evaluated the performance of different PROMs in measuring outcomes following lower extremity joint replacement surgery. Harris et al [39] identified the OKS and WOMAC as the best-performing PROMs specific to the lower extremity. The study assessed the validity, reproducibility, and acceptability of the scoring systems. Similarly, Alviar et al [40] in their review found OKS and WOMAC to be the best PROMs after assessing the quality of patient-reported outcome scoring systems. Collins and Roos [41] analyzed attributes of the 11 most frequently used PROMs for total hip replacement and TKA and considered KOOS, WOMAC, and OKS to be good PROMs. These PROMs aid in monitoring progress, guiding interventions, and facilitating shared decision-making. Nevertheless, there are limitations, such as the risk of subjective bias and variability in interpretation. Patient responses may be influenced by personal perceptions, mood, cognitive state, or cultural differences. Obtaining consistent and accurate data can be challenging, especially if patients struggle to recall specific details over time. To gain a more thorough understanding of the impact of TKA, it is crucial to integrate objective measures of muscle strength and functional assessments, particularly in addressing muscle weakness.

## Strength Measurements

The indirect assessment of muscle function in any of the PROMs may not provide sufficient indication of dysfunction, necessitating a comprehensive analysis of objective assessment tools in TKA (Table 1)(Multimedia Appendix 1). To this end, this section will focus on the types of muscle strength assessments that are feasible in a TKA clinical setting.

**Table .** Comparative analysis of outcome measures used in total knee arthroplasty (TKA), including patient-reported outcomes, muscle strength, muscle mass, and physical performance tools. Comparison criteria include measurement type, reliability, validity, ease of use, clinical relevance, limitations, and best use case.

| Comparison | PROMs[a] (KSS[b], WOMAC[c], KOOS[d], and OKS)[e] | Muscle strength tests (IKD[f], PFD[g], and HHD[h]) | Muscle mass tests (CT[i], MRI[j], DXA[k], and BIA[l]) | Physical performance tests (TUG[m], 6MWT[n], 5R-STS[o], and SCT[p]) |
|---|---|---|---|---|
| Measurement type | Subjective (patient-reported pain, function) | Objective (muscle force output) | Objective (muscle cross-sectional area or mass) | Objective (functional movement performance) |
| Reliability | High test-retest reliability but can be influenced by patient perception | High reliability but dependent on test consistency and patient effort | High for CT, MRI, and DXA; moderate for BIA (affected by hydration) | High for TUG, 6MWT, and 5R-STS |
| Validity | Valid for subjective function but may not correlate with actual muscle strength or mobility | IKD, highest validity. PFD, good validity with proper fixation. HHD, lower validity | Highly valid for assessing atrophy or hypertrophy | Strong correlation with mobility and functional independence |
| Ease of use | Simple, noninvasive, and patient-friendly | Requires equipment; portable dynamometers are easier than isokinetic ones | CT or MRI is expensive; DXA is more accessible; BIA is easiest | Quick, requires minimal equipment (eg, chair and stopwatch) |
| Clinical relevance | Useful for tracking patient-perceived recovery | Directly assesses quadriceps and hamstring strength, critical for TKA recovery | Helps detect muscle loss post-TKA, which can affect long-term function | Strong predictor of fall risk, mobility, and independence |
| Limitations | Subjective, may not reflect real functional capacity | Requires patient cooperation; costly for isokinetic dynamometers | Expensive (CT and MRI), radiation exposure (CT and DXA), less precise (BIA) | Can be influenced by patient motivation, fatigue, or comorbidities |
| Best use case | Tracking patient-perceived progress | Measuring post-TKA quadriceps or hamstring recovery | Evaluating long-term muscle loss and sarcopenia | Assessing mobility and real-world function |

[a]PROM: patient-reported outcome measure.

[b]KSS: Knee Society Score.

[c]WOMAC: Western Ontario and McMaster Universities Osteoarthritis Index.

[d]KOOS: Knee Injury and Osteoarthritis Outcome Score.

[e]OKS: Oxford Knee Score.

[f]IKD: isokinetic dynamometer.

[g]PFD: portable fixed dynamometer.

[h]HHD: handheld dynamometry.

[i]CT: computed tomography.

[j]MRI: magnetic resonance imaging.

[k]DXA: dual-energy X-ray absorptiometry.

[l]iBIA: bioelectrical impedance analysis.

[m]TUG: Time Up and Go.

[n]6MWT: 6-Minute Walk Test.

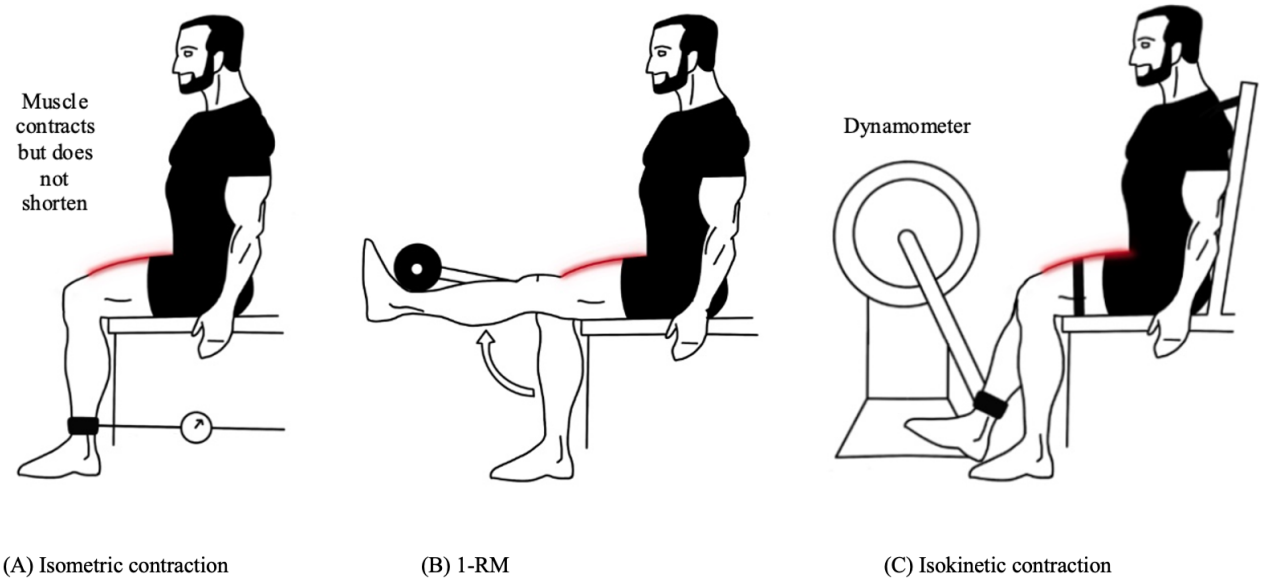
[o]5R-STS: 5-Repetition Sit-to-Stand.

[p]SCT: Stair Climb Test.

When evaluating muscle function, the measured outcomes commonly include muscle strength and power [42]. Various methodologies assess these parameters, ranging from manual muscle testing to more complex and expensive isokinetic assessments. However, no general consensus exists regarding the preferred method of assessing muscle function in orthopedic practice, as each approach has distinct advantages and limitations [43].

Isometric strength assessment measures the maximum force generated during a contraction in which the muscle length remains constant (Figure 1A). Although isometric contractions are relatively uncommon in daily activities, their clinical relevance lies in their ability to predict functional capacity, particularly in older individuals and those with significant functional impairments [44]. Furthermore, isometric strength assessments have demonstrated high reliability across different orthopedic populations [45,46]. Isometric strength assessments can be easily performed using a dynamometer or force plate, and the individual is instructed to push or pull against an immovable object. Portable fixed dynamometer (PFD) has demonstrated high reliability (intraclass correlation coefficient [ICC] >0.90) and validity in measuring strong muscle contractions such as knee extension [47-49]. Importantly, studies have shown that knee muscle strength measurements using a PFD highly correlate with those obtained from an IKD, making PFD a viable alternative for clinical assessment [50].

**Figure 1.** Illustration of 3 common muscle strength assessment methods used in total knee arthroplasty (TKA) recovery: (A) isometric contraction—muscle generates force without joint movement; (B) one-repetition maximum (1-RM)—maximal weight lifted once with proper form; and (C) isokinetic contraction—muscle contracts at a constant speed throughout the range of motion using a dynamometer.




Toonstra and Mattacola [51] compared the reliability of IKD, PFD, and handheld dynamometry for isometric knee strength assessment. Their findings revealed high test-retest reliability for both IKD and PFD, while handheld dynamometry showed fair to poor reliability.

The strength of IKD lies in its ability to provide stabilization during testing and standardized protocols, making it the gold standard. However, PFD demonstrated comparable reliability while offering greater portability, ease of use, and cost-effectiveness. This makes PFD an attractive option for routine clinical evaluations, particularly in settings where IKD is not feasible.

Another key advantage of isometric assessments is their safety. Unlike dynamic strength tests, which involve movement and may place stress on healing tissues, isometric assessments minimize joint strain. This makes them particularly suitable for early postoperative assessments, where patient safety is a priority. Given these benefits, isometric strength measurements serve as a practical and reliable method for monitoring recovery following TKA.

One-repetition maximum (1-RM): Muscle strength can also be assessed through the one–repetition maximum (1-RM) test, which determines the maximum load that can be lifted in a single attempt, such as during a squat or leg extension (Figure 1B). This method evaluates muscle strength in a more functional and dynamic context, offering insight into the ability to generate force during real-world movements.

Despite its functional relevance, 1-RM testing has limitations. The requirement for specialized equipment (eg, weights and resistance machines) and the time-intensive nature of testing make it less practical for large patient cohorts. Additionally, 1-RM testing carries a risk of injury, particularly in postoperative populations where patients may not yet be able to safely perform maximal-effort movements. To mitigate this risk, researchers often use submaximal testing protocols to estimate 1-RM values [52].

Isokinetic testing measures muscle strength as the peak torque generated during a contraction performed at a constant angular velocity (Figure 1C). This method provides resistance that matches the individual's effort throughout the ROM, allowing for maximal force production at different joint angles. Isokinetic assessments are particularly valuable for evaluating muscle imbalances, monitoring rehabilitation progress, and understanding strength deficits following TKA.

Despite these advantages, isokinetic testing has notable limitations. The high cost and lack of portability make it less accessible in many clinical settings [53,54]. Additionally, isokinetic assessments require specialized equipment and trained personnel, making them impractical for routine postoperative evaluations. The potential for joint irritation and discomfort during testing further limits its suitability for early rehabilitation stages.

A key consideration when choosing between isometric and isokinetic testing is clinical feasibility. While isokinetic assessments provide detailed torque-angle relationships, they are not always necessary for functional recovery monitoring. In contrast, isometric testing provides a simple, reliable, and cost-effective means of assessing muscle strength, making it better suited for routine evaluations.

Lauermann et al [55] compared quadriceps strength assessments using isometric, isokinetic, and 1-RM methods in patients who underwent TKA. All 3 methods showed similar validity in detecting strength deficits, suggesting that isometric testing is an effective alternative to more complex assessments. Similarly, Lienhard et al [45] found comparable test-retest reliability across

these methods in patients who underwent TKA, reinforcing the viability of isometric assessments in clinical practice.

### Muscle Mass Measurements

Studies have shown a significant correlation between skeletal muscle mass and cross-sectional skeletal muscle area in the extremities with muscle strength or power [56,57]. Hence, assessing these morphological features could be a good way to gauge muscle function. Furthermore, with certain imaging modalities, it may also be possible to determine the extent of sarcopenia, muscle wasting, and the level of fibrosis and fat infiltration—all factors that play a significant role in muscle function.

Computed tomography (CT) is an imaging modality that was introduced over 50 years ago, as the first clinically accepted tool for body composition measurement and served as a gold standard [58]. This method uses X-ray beams to create cross-sectional images of the body, allowing estimation of total body fat, visceral fat, and skeletal muscle mass [59,60]. Despite its early acceptance and high validity for assessing limb muscle cross-sectional area, with excellent test–retest and inter-observer reliability (ICC 0.98 - 1.00 for thigh muscle measurements), CT has general limitations such as high costs, the need for skilled technicians, and radiation exposure [61,62]. However, in current practice, CT is rarely used for measuring muscle mass, likely due to its associated limitations and the emergence of alternative, more practical techniques such as magnetic resonance imaging (MRI).

The introduction of MRI in the 1980s replaced the CT scan as the gold standard [63]. Its 3D images of skeletal muscle, fat tissue, and other organs are created by emitting radiofrequency energy from hydrogen atom nuclei in magnetic fields, with signal variations indicating different tissue types [63,64]. This development brought about high-resolution images without radiation exposure, making it suitable for tracking small changes over time, which is beneficial for intervention and observational studies. Advancements in MRI techniques have notably reduced image acquisition times, and modern scanners can accommodate obese individuals. Validated against direct anatomical measurements ($r$=0.97), MRI exhibits high test-retest and inter-observer reliability (2.9%, $r$=0.99% and 2.6%, $r$=0.99) in healthy populations [63,64]. However, limitations in clinical and research settings include high costs, the need for technical expertise, space requirements, infeasibility for patients with claustrophobia or those with MRI-incompatible implanted devices (eg, cardiac pacemakers), and standardization is hampered by the existence of various data acquisition protocols [61,65].

Dual-energy X-ray absorptiometry (DXA) has emerged as a strong alternative to the gold standard for assessing body composition, being relatively cheap compared with CT scan and MRI, and easy to perform. Initially designed for measuring bone mineral density, DXA is now widely used for examining overall body composition and muscle mass [63]. DXA operates by using 2 X-ray beams to differentiate between fat, bone, and lean tissue based on their X-ray absorption properties [66]. This technique has shown a high correlation with both MRI and CT in estimating skeletal muscle mass, indicating its reliability ($r$=0.88 and $r$=0.77 - 0.95, respectively) [67-69]. Additionally, DXA test-retest reliability for measures of fat-free mass demonstrated a high correlation ($r$=0.99), with low precision errors ranging from below 1% to 3% [70].

However, DXA does present some limitations. First, it does not directly quantify muscle mass; instead, it calculates lean soft tissue mass or fat-free mass based on the different gray tones observed in the DXA scan. Assumptions are made during this calculation, and factors like dehydration and edema, which can be common in populations with obesity, may interfere with the accuracy of the measurements. Additionally, the lack of standardization across devices, software packages, and versions can yield different results, affecting the reliability and comparability of measurements. Moreover, while DXA involves less radiation exposure compared to CT, there is still some exposure [71,72].

Bioelectrical impedance analysis (BIA) has become widely used for assessing body composition in both clinical and nonclinical settings [73]. BIA operates on the principle that tissues with higher water and electrolyte content, like skeletal muscle, offer less resistance to the passage of a low-voltage electrical current compared to lipid-rich adipose tissue, such as bone. This conductivity difference is exploited by BIA systems to quantify different body compartments. Measurements obtained along with parameters like sex, age, and body weight are then integrated into the population-specific body composition prediction equations [64,72]. The advantages of BIA stem from its noninvasive nature, cost-effectiveness, and ease of use, but limitations include sensitivity to factors like body position, physical exercise, food intake, hydration status, and the need for population-specific equations [74]. A study by Buckinx et al [75] attempted to gauge the reliability of BIA in assessing appendicular lean mass. The results showed high intraoperator reliability, indicated by an ICC of 0.89 when performed by the same operator, and interoperator reliability was also relatively high with an ICC of 0.77 when performed by different operators. However, the study revealed a notable discrepancy when comparing appendicular lean mass assessed by DXA and predicted by BIA, as evidenced by a low ICC of 0.37. Importantly, there exists a potential for a significant prediction error at the individual level with BIA, coupled with a systematic positive bias leading to an overall underestimation of lean body mass measurements [76]. Additionally, hydration status is a frequent confounder in clinical settings, as BIA relies on electrical conductivity that is highly sensitive to fluctuations in body water content, potentially compromising accuracy and limiting its reliability in certain patient populations. Despite these limitations, BIA remains a viable alternative in situations where more precise methods are not feasible.

### Performance Measurements

While measurements of muscle mass and function are important for understanding physiological dysfunction and pathology, assessing physical performance is essential. In this context, functional performance-based tests use an individual's body weight as resistance, quantifying performance by the time taken or the number of repetitions completed. Several authoritative groups, including The Osteoarthritis Research Society

International, Rehabilitative Care Alliance, and the American Physical Therapy Association, recommend performance-based measures for individuals with osteoarthritis and arthroplasty [77-79]. These measures typically include assessments of gait speed tests, Stair Climb Tests (SCT), and Sit-to-Stand (STS) tests, each with distinct yet overlapping focuses.

Gait speed tests primarily assess mobility, endurance, and functional independence by measuring the time required to walk a set distance. They are simple, reliable, and highly predictive of overall recovery, fall risk, and mortality [80]. SCTs focus on functional strength, coordination, and balance, as stair negotiation requires greater knee flexion and quadriceps activation than level walking [81]. SCTs are more sensitive to persistent functional deficits than gait speed tests, as stair climbing is often more challenging postoperatively. STS tests, on the other hand, primarily evaluate lower limb strength and endurance by measuring the ability to transition from sitting to standing multiple times within a set duration or repetition count. STS tests strongly correlate with quadriceps strength and are useful for monitoring functional recovery [82].

While all 3 measures provide valuable insights into TKA recovery, they differ in their primary focus and clinical applicability. Gait speed and SCTs assess mobility in dynamic movement tasks, while STS focuses on lower limb muscle strength in a controlled, stationary task. Gait speed tests are most effective for evaluating overall mobility and endurance, whereas SCTs are ideal for assessing functional strength and dynamic balance [83]. In contrast, STS tests specifically target muscle power and endurance without assessing walking ability or balance [84]. Despite these differences, all three tests share a common goal of evaluating lower limb function and have been shown to be reliable, valid, and responsive to post-TKA functional improvements [78]. However, it is worth noting that both gait speed and SCTs may present logistical challenges in smaller clinical settings due to space constraints and increased time requirements for setup and execution, potentially limiting their routine use in busy outpatient environments.

Dobson et al [78] put forth a set of recommended performance-based measures for individuals diagnosed with hip and knee osteoarthritis or following joint replacement. This selection was based on expert surveys and systematic reviews, considering the evidence supporting the measurement properties of the tests, their feasibility, scoring methods, and expert consensus. The resulting recommended measures included the 30-Second Chair Stand Test (30CST), 40 m Fast-paced Walk Test, SCT, Time Up and Go (TUG), and 6-Minute Walk Test (6MWT). In a subsequent update by Dobson et al [85], the 10 Meter Fast-paced Walk Test and 20-second stair climb test were suggested as alternatives due to complexities in administering the 40-meter Fast-paced Walk and scoring the 11 Step Stair Climb Test.

A task force led by Westby et al [86] created the Total Joint Arthroplasty and Outcome Measures toolkit to be used before and after arthroplasty, which included: 30-second STS, gait speed, stair climb test, single leg stance test, 6MWT, TUG, and functional reach. Prior to the creation of this toolkit, Zeni et al [87] developed the Delaware Osteoarthritis Profile, a comprehensive set of tests that have been effectively used to measure functional performance pre and post knee arthroplasty, which include: TUG, SCT, and 6MWT.

While the literature contains numerous tests, the focus here will be on muscle strength performance tests. STS with its versions has been identified by Bergquist et al [88] in their systematic literature review to be the most appropriate performance-based clinical muscle strength test.

The 30CST is a component of the Senior Fitness Test and involves counting the number of sit-to-stand repetitions completed within a 30-second timeframe. Individuals who are more than halfway through a repetition at the 30-second mark are credited with completing the final repetition. This test has shown good to excellent reliability, with ICCs of approximately 0.84 to 0.92 in clinical populations, and its validity is supported by moderate to strong correlations with leg press strength measures [89].

The 5-Repetition Sit-to-Stand (5R-STS) test has gained widespread use as either a component of the short physical performance battery or as a standalone assessment tool in numerous studies. This test involves performing 5 sit-to-stand repetitions from a standard chair (44 - 48 cm height), with timing commencing upon a specific command or the initiation of the first movement and ending when the fifth stand-up is accomplished or when the patient sits down following the fifth stand-up [90]. The reliability of the 5R-STS test has been estimated across 10 studies, yielding a coefficient of 0.81, as reported by Bohannon (2011) [84]. The validity of this test as a measure of lower limb functional muscle strength is supported by its strong correlation with knee extension strength, as demonstrated by Lord et al [82]. Furthermore, the correlation between performance on the 5R-STS test and the TUG test, as well as gait speed tests, adds further support for its validity, as highlighted by Schaubert and Bohannon [91].

Although the 5R-STS and 30CST tests involve identical movements, they are not interchangeable. The 5R-STS indicates lower limb speed and power for those with lower physical function post-TKA, while the 30CST measures lower limb endurance for those with higher physical function [92].

While outcome measures are often viewed as purely psychometric, their clinical relevance is established through their role as important drivers of long-term surgical success. For instance, a recent systematic review and meta-analysis by Sumbal et al [93] demonstrated that loss of muscle mass, or sarcopenia, is a significant risk factor for prosthetic loosening, which remains a primary indication for revision surgery. Furthermore, physical performance levels and quadriceps strength have been shown to be strongly associated with patient satisfaction and restoration of functional ability after TKA [94]. More recently, Akatsuka et al [95] reinforced this clinical relevance by reporting significant correlations between quadriceps muscle mass and postoperative satisfaction, suggesting that these objective measures are predictive of how a patient perceives their surgical result.

## Discussion

### Principal findings

A range of assessment tools is available to evaluate recovery following TKA, each with distinct advantages and limitations. This review highlights the use, strengths, and constraints of these measures and recommends a comprehensive, multidimensional evaluation framework. Such an approach enhances the accuracy and clinical relevance of recovery assessments by triangulating data across patient-reported outcomes, strength measures, functional performance, and muscle mass.

Among muscle mass assessment techniques, DXA remains the gold standard due to its high precision and reliability. However, alternative methods such as BIA offer a noninvasive, lower-cost option suitable for clinical settings, despite reduced accuracy and sensitivity to minor changes in muscle composition.

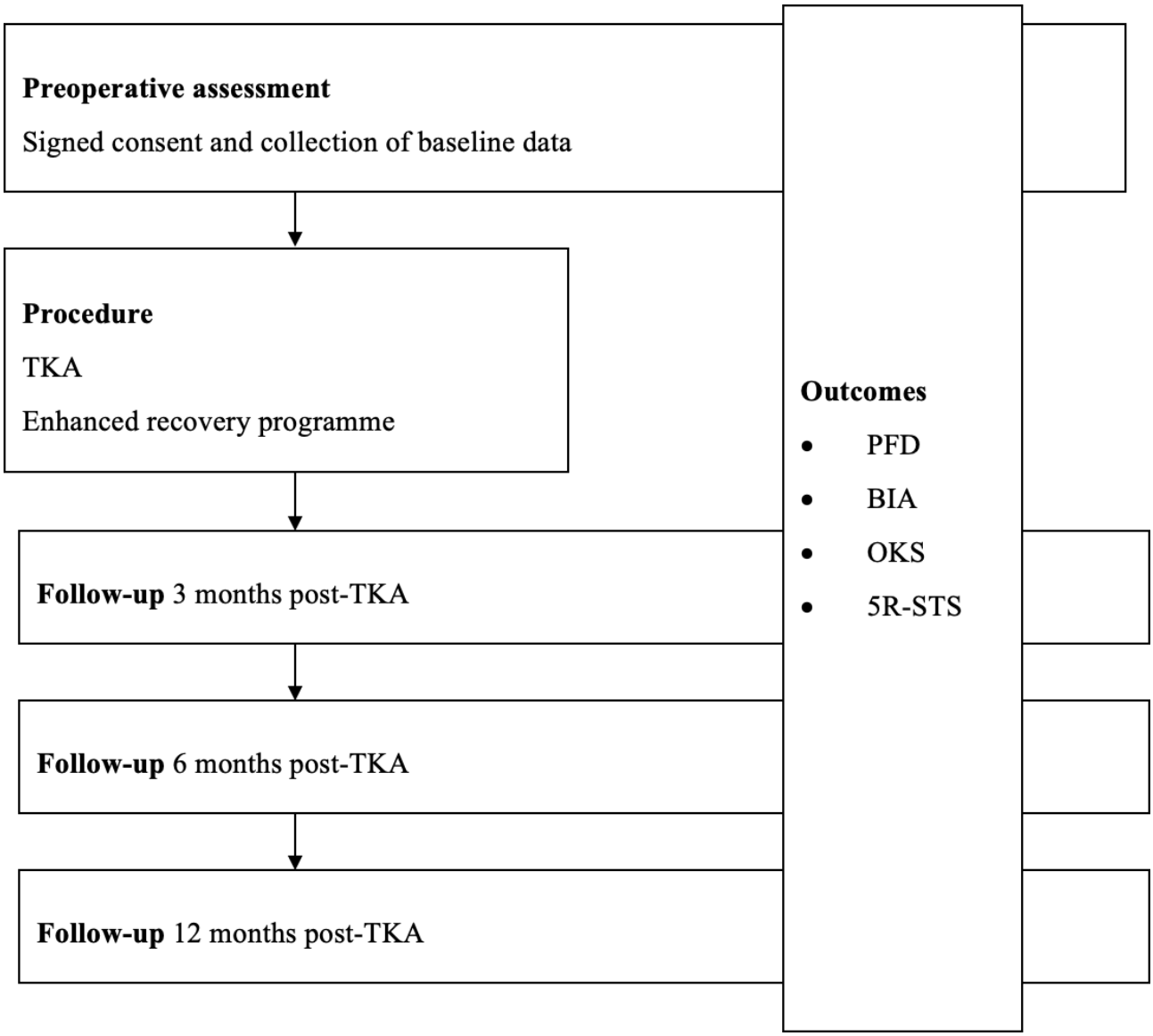
PROMs provide valuable insights into patient perspectives. The OKS is widely employed due to its ease of use, absence of ceiling or floor effects in the short term, and extensive validation in clinical research. KOOS, while broader in scope, often demonstrates ceiling effects and a higher completion burden. WOMAC overlaps conceptually but lacks OKS's responsiveness in early recovery. KSS combines patient and clinician input but is less feasible for routine use due to scoring complexity and reliance on in-person assessment. OKS offers the optimal balance of validity, efficiency, and clinical use in TKA follow-up. Nevertheless, PROMs alone do not provide objective strength assessments, which are critical for a comprehensive evaluation.

In contrast, IKD offers precise quantification of muscle strength, yet its high cost and operational complexity restrict its widespread application. As a practical alternative, PFD provides a reliable, cost-effective, and portable means of assessing muscle strength, making it highly suitable for routine clinical practice.

Performance-based measures such as the 5R-STS test offer objective, time-efficient insights into lower limb function. Compared to the TUG, 6MWT, or SCT, the 5R-STS requires less space, is less influenced by cardiovascular limitations, and is more feasible for patients with early post-op mobility impairments. It also shows strong correlation with knee extension strength and mobility metrics, enhancing its predictive use in post-TKA recovery.

This specific combination, OKS, PFD, BIA, and 5R-STS is what we have recommended in post-TKA follow-up (Figure 2) because it balances psychometric validity, clinical feasibility, and comprehensive recovery profiling. It provides a more complete view of TKA recovery than any individual tool. While alternatives like the 6MWT or SCT yield valuable data, their requirements for time, space, or greater cardiopulmonary reserve make them less practical for routine follow-up or for patients with limited mobility or comorbidities. Likewise, the TUG, though simple, primarily reflects balance and walking ability but lacks sensitivity to muscle power or patient satisfaction domains.

**Figure 2.** Clinical workflow algorithm for integrating recommended outcome measures throughout the total knee arthroplasty (TKA) pathway. The flowchart outlines the timing and selection of assessment tools from the preoperative assessment appointment to the final follow-up. 5R-STS: 5-Repetition Sit-to-Stand; BIA: bioelectrical impedance analysis; OKS: Oxford Knee Score; PFD: portable fixed dynamometer.




## Limitations

Finally, it is important to note that the interpretability and use of these tools may vary across populations. For instance, sarcopenia, obesity, or cardiopulmonary conditions may skew performance-based results or BIA readings, requiring clinicians to contextualize outcomes based on age, BMI, and comorbid burden. By triangulating across these assessments, clinicians can reduce bias from any single measure and develop a more individualized recovery evaluation.

## Conclusions

Throughout this paper, evidence suggests that incorporating muscle-based and performance-based measures alongside PROMs is essential for a comprehensive and clinically relevant assessment of TKA recovery. However, future research should ensure there are multicenter trials to validate the integration of the proposed assessments, as well as consensus agreement with clinicians (ie, through a Delphi approach) which translates into guidelines. Additionally, a greater understanding of intrinsic skeletal muscle (mal)adaptations following TKA may provide valuable insight into the mechanisms underlying persistent functional limitations.

## Authors' Contributions

Conceptualization: DP, AB
Formal analysis: AB
Investigation: AB
Methodology: AB, DP

Supervision: NM, DP
Visualization: AB
Writing – original draft: AB
Writing – review & writing: AB, VM
All authors approved the final version of the manuscript.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Comparative analysis of outcome measures used in total knee arthroplasty (TKA), including patient-reported outcomes, muscle strength, muscle mass, and physical performance tools. Comparison criteria include measurement type, reliability, validity, ease of use, clinical relevance, limitations, and best use case.
[DOCX File, 14 KB - periop_v9i1e84011_app1.docx ]

## References


1. Neogi T. The epidemiology and impact of pain in osteoarthritis. Osteoarthritis Cartilage 2013 Sep;21(9):1145-1153. [doi: 10.1016/j.joca.2013.03.018] [Medline: 23973124]
2. Vos T, Lim SS, Abbafati C. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 2020 Oct 17;396(10258):1204-1222. [doi: 10.1016/S0140-6736(20)30925-9] [Medline: 33069326]
3. Jüni P, Reichenbach S, Dieppe P. Osteoarthritis: Rational approach to treating the individual. Best Pract Res Clin Rheumatol 2006 Aug;20(4):721-740. [doi: 10.1016/j.berh.2006.05.002] [Medline: 16979535]
4. Rissolio L, Sabatini L, Risitano S, et al. Is it the surgeon, the patient, or the device? A comprehensive clinical and radiological evaluation of factors influencing patient satisfaction in 648 total knee arthroplasties. J Clin Med 2021 Jun 12;10(12):2599. [doi: 10.3390/jcm10122599] [Medline: 34204628]
5. Lewis GN, Rice DA, Rashid U, McNair PJ, Kluger MT, Somogyi AA. Trajectories of pain and function outcomes up to 5 to 8 years following total knee arthroplasty. J Arthroplasty 2023 Aug;38(8):1516-1521. [doi: 10.1016/j.arth.2023.02.020] [Medline: 36805116]
6. Tungtrongjit Y, Weingkum P, Saunkool P. The effect of preoperative quadriceps exercise on functional outcome after total knee arthroplasty. J Med Assoc Thai 2012 Oct;95 Suppl 10:S58-S66. [Medline: 23451440]
7. Vajapey SP, Morris J, Spitzer AI, Glassman AH, Greco NJ, Li M. Outcome reporting patterns in total knee arthroplasty: a systematic review. J Clin Orthop Trauma 2020 Jul;11(Suppl 4):S464-S471. [doi: 10.1016/j.jcot.2020.05.014] [Medline: 32774013]
8. Mizner RL, Petterson SC, Clements KE, Zeni JA, Irrgang JJ, Snyder-Mackler L. Measuring functional improvement after total knee arthroplasty requires both performance-based and patient-report assessments: a longitudinal analysis of outcomes. J Arthroplasty 2011 Aug;26(5):728-737. [doi: 10.1016/j.arth.2010.06.004] [Medline: 20851566]
9. Naili JE, Iversen MD, Esbjörnsson AC, et al. Deficits in functional performance and gait one year after total knee arthroplasty despite improved self-reported function. Knee Surg Sports Traumatol Arthrosc 2017 Nov;25(11):3378-3386. [doi: 10.1007/s00167-016-4234-7] [Medline: 27436196]
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. Insall JN, Dorr LD, Scott RD, Scott WN. Rationale of the Knee Society clinical rating system. Clin Orthop Relat Res 1989 Nov(248):13-14. [Medline: 2805470]
12. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988 Dec;15(12):1833-1840. [Medline: 3068365]
13. Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynnon BD. Knee Injury and Osteoarthritis Outcome Score (KOOS)—development of a self-administered outcome measure. J Orthop Sports Phys Ther 1998 Aug;28(2):88-96. [doi: 10.2519/jospt.1998.28.2.88] [Medline: 9699158]
14. Martimbianco ALC, Calabrese FR, Iha LAN, Petrilli M, Lira Neto O, Carneiro Filho M. Reliability of the “American Knee Society Score” (AKSS) [Article in Portuguese]. Acta Ortop Bras 2012;20(1):34-38. [doi: 10.1590/S1413-78522012000100007] [Medline: 24453578]
15. Hodges PW, Smeets RJ. Interaction between pain, movement, and physical activity: short-term benefits, long-term consequences, and targets for treatment. Clin J Pain 2015 Feb;31(2):97-107. [doi: 10.1097/AJP.0000000000000098] [Medline: 24709625]
16. Merkle SL, Sluka KA, Frey-Law LA. The interaction between pain and movement. J Hand Ther 2020;33(1):60-66. [doi: 10.1016/j.jht.2018.05.001] [Medline: 30025839]

17. Ghanem E, Pawasarat I, Lindsay A, et al. Limitations of the Knee Society Score in evaluating outcomes following revision total knee arthroplasty. J Bone Joint Surg Am 2010 Oct 20;92(14):2445-2451. [doi: 10.2106/JBJS.I.00252] [Medline: 20962195]
18. Vogel N, Kaelin R, Rychen T, Wendelspiess S, Müller-Gerbl M, Arnold MP. Comparison of responsiveness of patient-reported outcome measures after total knee arthroplasty. J Arthroplasty 2024 Jun;39(6):1487-1495.e2. [doi: 10.1016/j.arth.2023.11.029] [Medline: 38043745]
19. Bach CM, Nogler M, Steingruber IE, et al. Scoring systems in total knee arthroplasty. Clin Orthop Relat Res 2002 Jun;399(399):184-196. [doi: 10.1097/00003086-200206000-00022] [Medline: 12011708]
20. Karimijashni M, Abtahi F, Abbasalipour S, et al. Functional patient-reported outcome measures after hip or knee arthroplasty: a systematic review and content analysis using the International Classification of Functioning, Disability, and Health. Arthritis Care Res (Hoboken) 2024 Dec;76(12):1703-1722. [doi: 10.1002/acr.25413] [Medline: 39119739]
21. Noble PC, Scuderi GR, Brekke AC, et al. Development of a new Knee Society scoring system. Clin Orthop Relat Res 2012 Jan;470(1):20-32. [doi: 10.1007/s11999-011-2152-z] [Medline: 22065240]
22. Scuderi GR, Bourne RB, Noble PC, Benjamin JB, Lonner JH, Scott WN. The new Knee Society knee scoring system. Clin Orthop Relat Res 2012 Jan;470(1):3-19. [doi: 10.1007/s11999-011-2135-0] [Medline: 22045067]
23. Siljander MP, McQuivey KS, Fahs AM, Galasso LA, Serdahely KJ, Karadsheh MS. Current trends in patient-reported outcome measures in total joint arthroplasty: a study of 4 major orthopaedic journals. J Arthroplasty 2018 Nov;33(11):3416-3421. [doi: 10.1016/j.arth.2018.06.034] [Medline: 30057269]
24. Ritter MA, Davis KE, Davis P, et al. Preoperative malalignment increases risk of failure after total knee arthroplasty. J Bone Joint Surg Am 2013 Jan 16;95(2):126-131. [doi: 10.2106/JBJS.K.00607] [Medline: 23324959]
25. Capin JJ, Bade MJ, Jennings JM, Snyder-Mackler L, Stevens-Lapsley JE. Total knee arthroplasty assessments should include strength and performance-based functional tests to complement range-of-motion and patient-reported outcome measures. Phys Ther 2022 Jun 3;102(6):pzac033. [doi: 10.1093/ptj/pzac033] [Medline: 35358318]
26. Ben-Shlomo Y, Blom A, Boulton C, et al. The National Joint Registry 19th annual report 2022. : National Joint Registry; 2022 URL: https://www.njrcentre.org.uk/wp-content/uploads/NJR-Supported-Research.pdf [accessed 2025-03-18]
27. Lee MK, Naessens JM, Eton DT, O'Byrne TJ, Nyman MA. Functional outcomes and health-related quality of life before and after primary total knee replacement for patients from diverse geographic regions. J Arthroplasty 2021 Jun;36(6):1958-1965. [doi: 10.1016/j.arth.2021.01.043] [Medline: 33581972]
28. Beard DJ, Knezevic K, Al-Ali S, Dawson J, Price AJ. The use of outcome measures relating to the knee. Orthop Trauma 2010 Aug;24(4):309-316. [doi: 10.1016/j.mporth.2010.04.001]
29. Harris K, Lim CR, Dawson J, Fitzpatrick R, Beard DJ, Price AJ. The Oxford Knee Score and its subscales do not exhibit a ceiling or a floor effect in knee arthroplasty patients: an analysis of the National Health Service PROMs data set. Knee Surg Sports Traumatol Arthrosc 2017 Sep;25(9):2736-2742. [doi: 10.1007/s00167-015-3788-0] [Medline: 26519187]
30. Marx RG, Jones EC, Atwan NC, Closkey RF, Salvati EA, Sculco TP. Measuring improvement following total hip and knee arthroplasty using patient-based measures of outcome. J Bone Joint Surg Am 2005 Sep;87(9):1999-2005. [doi: 10.2106/JBJS.D.02286] [Medline: 16140815]
31. Beard DJ, Davies LJ, Cook JA, et al. The clinical and cost-effectiveness of total versus partial knee replacement in patients with medial compartment osteoarthritis (TOPKAT): 5-year outcomes of a randomised controlled trial. Lancet 2019 Aug 31;394(10200):746-756. [doi: 10.1016/S0140-6736(19)31281-4] [Medline: 31326135]
32. Escobar A, Gonzalez M, Quintana JM, et al. Patient acceptable symptom state and OMERACT-OARSI set of responder criteria in joint replacement. Identification of cut-off values. Osteoarthritis Cartilage 2012 Feb;20(2):87-92. [doi: 10.1016/j.joca.2011.11.007] [Medline: 22155074]
33. Berth A, Urbach D, Awiszus F. Improvement of voluntary quadriceps muscle activation after total knee arthroplasty. Arch Phys Med Rehabil 2002 Oct;83(10):1432-1436. [doi: 10.1053/apmr.2002.34829] [Medline: 12370881]
34. Bellamy N. WOMAC: A 20-year experiential review of a patient-centered self-reported health status questionnaire. J Rheumatol 2002 Dec;29(12):2473-2476. [Medline: 12465137]
35. Roos EM, Toksvig-Larsen S. Knee injury and Osteoarthritis Outcome Score (KOOS)—validation and comparison to the WOMAC in total knee replacement. Health Qual Life Outcomes 2003 May 25;1(1):17. [doi: 10.1186/1477-7525-1-17] [Medline: 12801417]
36. Antosh IJ, Svoboda SJ, Peck KY, Garcia EJ, Cameron KL. Change in KOOS and WOMAC scores in a young athletic population with and without anterior cruciate ligament injury. Am J Sports Med 2018 Jun;46(7):1606-1616. [doi: 10.1177/0363546518768753] [Medline: 29733680]
37. Collins NJ, Prinsen CAC, Christensen R, Bartels EM, Terwee CB, Roos EM. Knee injury and Osteoarthritis Outcome Score (KOOS): systematic review and meta-analysis of measurement properties. Osteoarthritis Cartilage 2016 Aug;24(8):1317-1329. [doi: 10.1016/j.joca.2016.03.010] [Medline: 27012756]
38. Wang Y, Yin M, Zhu S, Chen X, Zhou H, Qian W. Patient-reported outcome measures used in patients undergoing total knee arthroplasty. Bone Joint Res 2021 Mar;10(3):203-217. [doi: 10.1302/2046-3758.103.BJR-2020-0268.R1] [Medline: 33734821]

39. Harris K, Dawson J, Gibbons E, et al. Systematic review of measurement properties of patient-reported outcome measures used in patients undergoing hip and knee arthroplasty. Patient Relat Outcome Meas 2016;7:101-108. [doi: 10.2147/PROM.S97774] [Medline: 27524925]
40. Alviar MJ, Olver J, Brand C, et al. Do patient-reported outcome measures in hip and knee arthroplasty rehabilitation have robust measurement attributes? A systematic review. J Rehabil Med 2011 Jun;43(7):572-583. [doi: 10.2340/16501977-0828] [Medline: 21607295]
41. 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]
42. Gerrits KH, Beltman MJ, Koppe PA, et al. Isometric muscle function of knee extensors and the relation with functional performance in patients with stroke. Arch Phys Med Rehabil 2009 Mar;90(3):480-487. [doi: 10.1016/j.apmr.2008.09.562] [Medline: 19254615]
43. Maffiuletti NA. Assessment of hip and knee muscle function in orthopaedic practice and research. J Bone Joint Surg Am 2010 Jan;92(1):220-229. [doi: 10.2106/JBJS.I.00305] [Medline: 20048117]
44. Hyun CW, Han EY, Im SH, et al. Hemiparetic knee extensor strength and balance function are predictors of ambulatory function in subacute stroke patients. Ann Rehabil Med 2015 Aug;39(4):577-585. [doi: 10.5535/arm.2015.39.4.577] [Medline: 26361594]
45. Norris R, Morrison S, Price A, et al. Inline dynamometry provides reliable measurements of quadriceps strength in healthy and ACL-reconstructed individuals and is a valid substitute for isometric electromechanical dynamometry following ACL reconstruction. Knee 2024 Jan;46:136-147. [doi: 10.1016/j.knee.2023.12.006] [Medline: 38142660]
46. Lienhard K, Lauermann SP, Schneider D, Item-Glatthorn JF, Casartelli NC, Maffiuletti NA. Validity and reliability of isometric, isokinetic and isoinertial modalities for the assessment of quadriceps muscle strength in patients with total knee arthroplasty. J Electromyogr Kinesiol 2013 Dec;23(6):1283-1288. [doi: 10.1016/j.jelekin.2013.09.004] [Medline: 24113423]
47. Ieiri A, Tushima E, Ishida K, Inoue M, Kanno T, Masuda T. Reliability of measurements of hip abduction strength obtained with a hand-held dynamometer. Physiother Theory Pract 2015 Feb;31(2):146-152. [doi: 10.3109/09593985.2014.960539] [Medline: 25264015]
48. Bohannon RW, Kindig J, Sabo G, Duni AE, Cram P. Isometric knee extension force measured using a handheld dynamometer with and without belt-stabilization. Physiother Theory Pract 2012 Oct;28(7):562-568. [doi: 10.3109/09593985.2011.640385] [Medline: 22191509]
49. Bohannon RW, Pritchard RO, Glenney SS. Portable belt–stabilized hand-held dynamometry set-up for measuring knee extension force. Isokinet Exerc Sci 2013 Nov;21(4):325-329. [doi: 10.3233/IES-130517]
50. Martins J, da Silva JR, da Silva MRB, Bevilaqua-Grossi D. Reliability and validity of the belt-stabilized handheld dynamometer in hip- and knee-strength Tests. J Athl Train 2017 Sep;52(9):809-819. [doi: 10.4085/1062-6050-52.6.04] [Medline: 28787180]
51. Toonstra J, Mattacola CG. Test-retest reliability and validity of isometric knee-flexion and -extension measurement using 3 methods of assessing muscle strength. J Sport Rehabil 2013 Feb 1;22(1):10.1123/jsr.2013.TR7. [doi: 10.1123/jsr.2013.TR7] [Medline: 22951307]
52. Abdalla PP, Carvalho ADS, Dos Santos AP, et al. One-repetition submaximal protocol to measure knee extensor muscle strength among older adults with and without sarcopenia: a validation study. BMC Sports Sci Med Rehabil 2020;12(1):29. [doi: 10.1186/s13102-020-00178-9] [Medline: 32391159]
53. Horstmann H, Medico P, Lasch F, Krutsch W, Weber-Spickschen TS. Simplified measurement of maximum strength after knee surgery: application-based knee-training device compared to isokinetic testing. Open Access J Sports Med 2020;11:43-49. [doi: 10.2147/OAJSM.S214598] [Medline: 32158284]
54. Whiteley R, Jacobsen P, Prior S, Skazalski C, Otten R, Johnson A. Correlation of isokinetic and novel hand-held dynamometry measures of knee flexion and extension strength testing. J Sci Med Sport 2012 Sep;15(5):444-450. [doi: 10.1016/j.jsams.2012.01.003] [Medline: 22424705]
55. Lauermann SP, Lienhard K, Item-Glatthorn JF, Casartelli NC, Maffiuletti NA. Assessment of quadriceps muscle weakness in patients after total knee arthroplasty and total hip arthroplasty: methodological issues. J Electromyogr Kinesiol 2014 Apr;24(2):285-291. [doi: 10.1016/j.jelekin.2013.10.018] [Medline: 24290027]
56. Tsukasaki K, Matsui Y, Arai H, et al. Association of muscle strength and gait speed with cross-sectional muscle area determined by mid-thigh computed tomography—a comparison with skeletal muscle mass measured by dual-energy X-ray absorptiometry. J Frailty Aging 2020;9(2):82-89. [doi: 10.14283/jfa.2020.16] [Medline: 32259181]
57. Riviati N, Indra B. Relationship between muscle mass and muscle strength with physical performance in older adults: a systematic review. SAGE Open Med 2023;11:20503121231214650. [doi: 10.1177/20503121231214650] [Medline: 38033420]
58. Heymsfield SB, Adamek M, Gonzalez MC, Jia G, Thomas DM. Assessing skeletal muscle mass: historical overview and state of the art. J Cachexia Sarcopenia Muscle 2014 Mar;5(1):9-18. [doi: 10.1007/s13539-014-0130-5] [Medline: 24532493]
59. Binay Safer V, Safer U. Usefulness and limitations of single-slice computed tomography analysis at the third lumbar region in the assessment of sarcopenia. Crit Care 2013 Nov 20;17(6):466. [doi: 10.1186/cc13123] [Medline: 24252497]

60. Gibson DJ, Burden ST, Strauss BJ, Todd C, Lal S. The role of computed tomography in evaluating body composition and the influence of reduced muscle mass on clinical outcome in abdominal malignancy: a systematic review. Eur J Clin Nutr 2015 Oct;69(10):1079-1086. [doi: 10.1038/ejcn.2015.32] [Medline: 25782424]
61. Lemos T, Gallagher D. Current body composition measurement techniques. Curr Opin Endocrinol Diabetes Obes 2017 Oct;24(5):310-314. [doi: 10.1097/MED.0000000000000360] [Medline: 28696961]
62. Strandberg S, Wretling ML, Wredmark T, Shalabi A. Reliability of computed tomography measurements in assessment of thigh muscle cross-sectional area and attenuation. BMC Med Imaging 2010 Aug 11;10(1):18. [doi: 10.1186/1471-2342-10-18] [Medline: 20701775]
63. Erlandson MC, Lorbergs AL, Mathur S, Cheung AM. Muscle analysis using pQCT, DXA and MRI. Eur J Radiol 2016 Aug;85(8):1505-1511. [doi: 10.1016/j.ejrad.2016.03.001] [Medline: 27005009]
64. Heymsfield SB, Gonzalez MC, Lu J, Jia G, Zheng J. Skeletal muscle mass and quality: evolution of modern measurement concepts in the context of sarcopenia. Proc Nutr Soc 2015 Nov;74(4):355-366. [doi: 10.1017/S0029665115000129] [Medline: 25851205]
65. Prado CMM, Heymsfield SB. Lean tissue imaging: a new era for nutritional assessment and intervention. JPEN J Parenter Enteral Nutr 2014 Nov;38(8):940-953. [doi: 10.1177/0148607114550189] [Medline: 25239112]
66. Lustgarten MS, Fielding RA. Assessment of analytical methods used to measure changes in body composition in the elderly and recommendations for their use in phase II clinical trials. J Nutr Health Aging 2011 May;15(5):368-375. [doi: 10.1007/s12603-011-0049-x] [Medline: 21528163]
67. Maden-Wilkinson TM, Degens H, Jones DA, McPhee JS. Comparison of MRI and DXA to measure muscle size and age-related atrophy in thigh muscles. J Musculoskelet Neuronal Interact 2013 Sep;13(3):320-328. [Medline: 23989253]
68. Kim J, Wang Z, Heymsfield SB, Baumgartner RN, Gallagher D. Total-body skeletal muscle mass: estimation by a new dual-energy X-ray absorptiometry method. Am J Clin Nutr 2002 Aug;76(2):378-383. [doi: 10.1093/ajcn/76.2.378] [Medline: 12145010]
69. Taylor AE, Kuper H, Varma RD, et al. Validation of dual energy X-ray absorptiometry measures of abdominal fat by comparison with magnetic resonance imaging in an Indian population. PLoS One 2012;7(12):e51042. [doi: 10.1371/journal.pone.0051042] [Medline: 23272086]
70. Hangartner TN, Warner S, Braillon P, Jankowski L, Shepherd J. The official positions of the International Society for Clinical Densitometry: acquisition of dual-energy X-ray absorptiometry body composition and considerations regarding analysis and repeatability of measures. J Clin Densitom 2013;16(4):520-536. [doi: 10.1016/j.jocd.2013.08.007] [Medline: 24183641]
71. Toombs RJ, Ducher G, Shepherd JA, De Souza MJ. The impact of recent technological advances on the trueness and precision of DXA to assess body composition. Obesity (Silver Spring) 2012 Jan;20(1):30-39. [doi: 10.1038/oby.2011.211] [Medline: 21760631]
72. Johnson Stoklossa CA, Forhan M, Padwal RS, Gonzalez MC, Prado CM. Practical considerations for body composition assessment of adults with class ii/iii obesity using bioelectrical impedance analysis or dual-energy x-ray absorptiometry. Curr Obes Rep 2016 Dec;5(4):389-396. [doi: 10.1007/s13679-016-0228-5] [Medline: 27591783]
73. Khalil SF, Mohktar MS, Ibrahim F. The theory and fundamentals of bioimpedance analysis in clinical status monitoring and diagnosis of diseases. Sensors (Basel) 2014 Jun 19;14(6):10895-10928. [doi: 10.3390/s140610895] [Medline: 24949644]
74. Marra M, Sammarco R, De Lorenzo A, et al. Assessment of body composition in health and disease using bioelectrical impedance analysis (BIA) and dual energy x-ray absorptiometry (DXA): a critical overview. Contrast Media Mol Imaging 2019;2019:3548284. [doi: 10.1155/2019/3548284] [Medline: 31275083]
75. Buckinx F, Reginster JY, Dardenne N, et al. Concordance between muscle mass assessed by bioelectrical impedance analysis and by dual energy x-ray absorptiometry: a cross-sectional study. BMC Musculoskelet Disord 2015 Mar 18;16(1):60. [doi: 10.1186/s12891-015-0510-9] [Medline: 25887598]
76. Ling CHY, de Craen AJM, Slagboom PE, et al. Accuracy of direct segmental multi-frequency bioimpedance analysis in the assessment of total body and segmental body composition in middle-aged adult population. Clin Nutr 2011 Oct;30(5):610-615. [doi: 10.1016/j.clnu.2011.04.001] [Medline: 21555168]
77. Jette DU, Hunter SJ, Burkett L, et al. Physical therapist management of total knee arthroplasty. Phys Ther 2020 Aug 31;100(9):1603-1631. [doi: 10.1093/ptj/pzaa099] [Medline: 32542403]
78. Dobson F, Hinman RS, Roos EM, et al. OARSI recommended performance-based tests to assess physical function in people diagnosed with hip or knee osteoarthritis. Osteoarthritis Cartilage 2013 Aug;21(8):1042-1052. [doi: 10.1016/j.joca.2013.05.002] [Medline: 23680877]
79. Rehabilitative care best practices for patients with primary hip & knee replacements. : Rehabilitative Care Alliance; 2017 URL: https://rehabcarealliance.ca/wp-content/uploads/2025/03/TJR_Framework.pdf [accessed 2026-02-13]
80. Studenski S, Perera S, Patel K, et al. Gait speed and survival in older adults. JAMA 2011 Jan 5;305(1):50-58. [doi: 10.1001/jama.2010.1923] [Medline: 21205966]
81. Fenner V, Behrend H, Kuster MS. Whole body gait function during stair ascending and level walking in patients following total knee arthroplasty. Int J Phys Med Rehabil 2014;02(5). [doi: 10.4172/2329-9096.1000231]

82. Lord SR, Murray SM, Chapman K, Munro B, Tiedemann A. Sit-to-stand performance depends on sensation, speed, balance, and psychological status in addition to strength in older people. J Gerontol A Biol Sci Med Sci 2002 Aug;57(8):M539-M543. [doi: 10.1093/gerona/57.8.m539] [Medline: 12145369]
83. Mizner RL, Petterson SC, Snyder-Mackler L. Quadriceps strength and the time course of functional recovery after total knee arthroplasty. J Orthop Sports Phys Ther 2005 Jul;35(7):424-436. [doi: 10.2519/jospt.2005.35.7.424] [Medline: 16108583]
84. Bohannon RW. Test-retest reliability of the five-repetition sit-to-stand test: a systematic review of the literature involving adults. J Strength Cond Res 2011 Nov;25(11):3205-3207. [doi: 10.1519/JSC.0b013e318234e59f] [Medline: 21904240]
85. Dobson F, Hinman RS, Hall M, et al. Reliability and measurement error of the Osteoarthritis Research Society International (OARSI) recommended performance-based tests of physical function in people with hip and knee osteoarthritis. Osteoarthritis Cartilage 2017 Nov;25(11):1792-1796. [doi: 10.1016/j.joca.2017.06.006] [Medline: 28647467]
86. Hoens A, Westby M, Longstaff S, Duggan M. Total Joint Arthroplasty and Outcome Measures (TJAOM) toolkit. The University of British Columbia. 2016. URL: https://physicaltherapy.med.ubc.ca/physical-therapy-knowledge-broker/tjaom-toolkit/ [accessed 2026-02-13]
87. Zeni JA, Axe MJ, Snyder-Mackler L. Clinical predictors of elective total joint replacement in persons with end-stage knee osteoarthritis. BMC Musculoskelet Disord 2010 May 6;11(1):86. [doi: 10.1186/1471-2474-11-86] [Medline: 20459622]
88. Bergquist R, Weber M, Schwenk M, et al. Performance-based clinical tests of balance and muscle strength used in young seniors: a systematic literature review. BMC Geriatr 2019 Jan 9;19(1):9. [doi: 10.1186/s12877-018-1011-0] [Medline: 30626340]
89. Gill S, Hely R, Page RS, Hely A, Harrison B, Landers S. Thirty second chair stand test: test-retest reliability, agreement and minimum detectable change in people with early-stage knee osteoarthritis. Physiother Res Int 2022 Jul;27(3):e1957. [doi: 10.1002/pri.1957] [Medline: 35592902]
90. Klukowska AM, Staartjes VE, Vandertop WP, Schröder ML. Five-repetition sit-to-stand test performance in healthy individuals: reference values and predictors from 2 prospective cohorts. Neurospine 2021 Dec;18(4):760-769. [doi: 10.14245/ns.2142750.375] [Medline: 35000330]
91. Schaubert KL, Bohannon RW. Reliability and validity of three strength measures obtained from community-dwelling elderly persons. J Strength Cond Res 2005 Aug;19(3):717-720. [doi: 10.1519/R-15954.1] [Medline: 16095431]
92. McCarthy EK, Horvat MA, Holtsberg PA, Wisenbaker JM. Repeated chair stands as a measure of lower limb strength in sexagenarian women. J Gerontol A Biol Sci Med Sci 2004 Nov;59(11):1207-1212. [doi: 10.1093/gerona/59.11.1207] [Medline: 15602077]
93. Sumbal R, Abbas M, Sheikh SM, Sumbal A. Prevalence and clinical impact of sarcopenia in patients undergoing total joint arthroplasty: a systematic review and a meta-analysis. J Arthroplasty 2024 Dec;39(12):3128-3135. [doi: 10.1016/j.arth.2024.06.021] [Medline: 38901711]
94. Furu M, Ito H, Nishikawa T, et al. Quadriceps strength affects patient satisfaction after total knee arthroplasty. J Orthop Sci 2016 Jan;21(1):38-43. [doi: 10.1016/j.jos.2015.10.002] [Medline: 26755384]
95. Akatsuka Y, Teramoto A, Takashima H, Okada Y, Watanabe K, Yamashita T. Relationships of cross-sectional area of the thigh muscles before or after total knee arthroplasty with postoperative pain or patient satisfaction: a retrospective, exploratory study. Asia Pac J Sports Med Arthrosc Rehabil Technol 2023 Jul;33:20-24. [doi: 10.1016/j.asmart.2023.08.006] [Medline: 37680195]


## Abbreviations

**1-RM:** 1–repetition maximum
**30CST:** 30-Second Chair Stand Test
**5R-STS:** 5-Repetition Sit-to-Stand
**6MWT:** 6-Minute Walk Test
**ADL:** activities of daily living
**BIA:** bioelectrical impedance analysis
**COSMIN:** Consensus-Based Standards for the Selection of Health Measurement Instruments
**CT:** computed tomography
**DXA:** dual-energy X-ray absorptiometry
**ICC:** intraclass correlation coefficient
**KOOS:** Knee Injury and Osteoarthritis Outcome Score
**KSS:** Knee Society Clinical Rating System
**MRI:** magnetic resonance imaging
**OKS:** Oxford Knee Score
**PFD:** portable fixed dynamometer
**PROM:** patient-related outcome measure
**ROM:** range of motion

**SCT:** Stair Climb Test
**STS:** Sit-to-Stand
**TKA:** total knee arthroplasty
**TUG:** Time Up and Go
**WOMAC:** Western Ontario and McMaster Universities Osteoarthritis

*Edited by N Rohatgi; submitted 12.Sep.2025; peer-reviewed by J Most; revised version received 30.Jan.2026; accepted 04.Feb.2026; published 25.Feb.2026.*

*Please cite as:*
*Boukabache A, Maruthainar N, Manhas V, Player D*
*Multidimensional Assessment of Recovery After Total Knee Arthroplasty in Clinical Practice: Critical Narrative Review*
*JMIR Perioper Med 2026;9:e84011*
*URL: https://periop.jmir.org/2026/1/e84011*
*doi:10.2196/84011*

# Forced-Air Warming Temperature Settings for Treating Postoperative Hypothermia in the Postanesthesia Care Unit: Randomized Controlled Trial

Koravee Pasutharnchat, MD; Rattaphol Seangrung, MD; Sirikarn Sirisophaphong, MD; Wilailuck Wongkum, BNS

Department of Anesthesiology, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, 270 Rama 6 Road, Ratchathewi, Bangkok, Thailand

**Corresponding Author:**
Koravee Pasutharnchat, MD
Department of Anesthesiology, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, 270 Rama 6 Road, Ratchathewi, Bangkok, Thailand

## Abstract

**Background:** Hypothermia, defined as a core body temperature below 36 °C, is a common postoperative complication associated with adverse outcomes, including delayed wound healing, infections, and increased bleeding.

**Objective:** This randomized controlled trial evaluated the efficacy of different forced-air warming system temperature settings in treating postoperative hypothermia in the postanesthesia care unit.

**Methods:** A total of 132 patients undergoing elective surgery at Ramathibodi Hospital between April 2023 and May 2024 were randomized into 3 groups (n=44 per group): group C (warming set to 38 °C), group F1 (warming set to 42 °C), and group F2 (warming set to 42 °C, reduced to 38 °C after achieving 36 °C). Tympanic temperature was recorded at 5-minute intervals during rewarming and every 10 minutes after normothermia (≥36 °C) was achieved. The primary outcome was rewarming time. Secondary outcomes included the incidence of temperature drops, hemodynamic parameters, adverse events, and patient comfort scores.

**Results:** Baseline characteristics and clinical variables, including vital signs, were comparable among groups ($P$>.05). Group F2 achieved the shortest mean rewarming time of 33.3 (SD 13.81) min; however, differences between groups were not statistically significant ($P$=.460). Group F2 had the lowest incidence of temperature drops below 36 °C after normothermia (1/44, 2.27%; $P$=.009). Group C had the highest incidence of rewarming exceeding 1 hour (10/44, 22.73%; $P$=.017).

**Conclusions:** While rewarming times were similar across groups, the protocol using an initial setting of 42 °C followed by a reduction to 38 °C (group F2) effectively minimized temperature drops after normothermia, suggesting its superiority for managing postoperative hypothermia in the postanesthesia care unit.

**Trial Registration:** Thaiclinicaltrials.org TCTR20231012004; https://www.thaiclinicaltrials.org/show/TCTR20231012004

***(JMIR Perioper Med 2026;9:e85045)*** doi:10.2196/85045

**KEYWORDS**

postoperative hypothermia; forced-air warming; effectiveness; postanesthesia care unit; temperature settings

## Introduction

Hypothermia, defined as a core body temperature below 36 °C, is a frequent complication in patients undergoing elective surgery [1]. Intraoperative hypothermia, if uncorrected, often leads to postoperative hypothermia, particularly in the recovery room, where insufficient warming measures can exacerbate the condition. The prevalence of postoperative hypothermia in the postanesthesia care unit (PACU) has been reported to range from 20% to 28% at arrival and from 18.5% to 26% within 30 minutes after arrival [2]. Postoperative hypothermia is clinically significant, as it has been associated with impaired wound healing, increased risk of surgical site infection, greater blood loss, cardiac arrhythmias, and prolonged hospitalization [3,4]. These adverse consequences highlight the importance of effective temperature management strategies throughout the perioperative period. Recent guidelines and reviews, including Enhanced Recovery After Surgery pathways and the clinical recommendations from the Royal College of Anesthesiologists of Thailand, emphasize the critical role of maintaining normothermia as a core component to reduce surgical site infections and hospital stay [5-8].

Active warming techniques, particularly forced-air warming (FAW), are widely implemented to reduce the incidence of perioperative hypothermia. FAW devices deliver warmed air (32 °C - 47 °C) through a specialized blanket, with built-in safety mechanisms to prevent overheating [9,10]. Systematic reviews have demonstrated that FAW is superior to conventional blankets, reducing the time to restore normothermia by more than an hour [11]. While these findings confirm its effectiveness in facilitating rewarming, the literature remains inconclusive

regarding the optimal temperature setting for postoperative use. Most previous studies have focused on preoperative or intraoperative warming [12-18], whereas evidence for postoperative FAW application remains limited. Xu et al [19] reported that FAW at 42 °C was more effective than at 38 °C or conventional blankets in elderly patients undergoing joint replacement. However, the generalizability of that study was restricted by the narrow patient population, short operative times, and limited assessment of adverse events.

At our institution, the prevalence of postoperative hypothermia has remained notable despite the routine availability of FAW systems. Pisitsak et al [20] documented hypothermia in 20% of patients under regional anesthesia and in 16% under general anesthesia. More recent institutional data from 2019 to 2022 indicate an incidence of 23% among surgical patients recovering in the PACU. Furthermore, between 2022 and 2024, the prevalence of hypothermia ranged from 10.8% to 13.8% despite widespread FAW use across multiple surgical specialties, including general surgery, orthopedics, otolaryngology, obstetrics and gynecology, and cardiac surgery (Department of Anesthesiology, Faculty of Medicine Ramathibodi Hospital, Mahidol University. Internal statistical data analyzed via Power BI dashboard, unpublished data, January 2025). These findings suggest that, in addition to patient- and procedure-related factors, variability in FAW temperature settings contributes to inconsistent outcomes.

Current practice in our PACUs uses FAW with adjustable temperature settings ranging from 38 °C to 42 °C; however, no standardized protocol exists to guide optimal temperature selection. This variability reflects broader uncertainty regarding the most effective strategy for postoperative rewarming and underscores the need for evidence-based guidance. To our knowledge, no prior randomized trial has evaluated a step-down temperature protocol (42 °C to 38 °C) in a mixed adult surgical population. By addressing this gap, the present study examines the effectiveness of different FAW temperature settings to inform a pragmatic and standardized PACU warming approach, with the goal of improving consistency in clinical practice and enhancing patient safety.

## Methods

### Study Design

This study was designed as a prospective randomized controlled trial.

### Patients

A total of 132 patients scheduled for elective surgery across various specialties, including general surgery, orthopedics, urology, otolaryngology, obstetrics and gynecology, and cardiac surgery, were enrolled between April 2023 and May 2024. The inclusion criteria consisted of patients aged 18 to 80 years, American Society of Anesthesiologists (ASA) physical status I to III, who were undergoing elective procedures under either general or regional anesthesia, with an expected operating time of at least 2 hours.

Exclusion criteria included patients with a core temperature exceeding 37.5 °C, evidence of infection (eg, sepsis), conditions precluding the use of forced-air warming (eg, burns, agitation, or delirium), those unable to communicate or complete the trial questionnaire, and patients who declined participation.

### Sample Size Calculation

A priori sample size calculation was conducted to ensure adequate statistical power for the study's primary outcome: the duration of forced-air warming required for a patient's core temperature to reach ≥36 °C. Based on a previous randomized controlled trial by Xu et al [19], utilizing a 2-sided significance level ($\alpha$=.05), adjusted for multiple comparisons among the 3 groups ($\alpha$/3=.017), corresponding to a $z$ score of 2.41, a statistical power of 80% ($z$ for $\beta$=0.84), an estimated SD of 6.45 minutes, and a clinically meaningful difference in rewarming time of 5 minutes, the calculation determined that 36 participants were required per group. To accommodate an anticipated 20% participant dropout rate, the sample size was prudently inflated to 44 participants for each of the 3 intervention groups. This resulted in a total sample size of 132 participants, ensuring robust statistical inference for our findings.

### Randomization

Randomization was performed using stratified block randomization with proportional allocation based on the type of anesthesia (general vs regional) to ensure balanced distribution of thermoregulatory impairment mechanisms across groups. A research assistant not involved in patient recruitment generated the computer-based random sequence using permuted blocks of variable size. Allocation was concealed using sequentially numbered, sealed opaque envelopes, which were opened only after participant enrollment. The study personnel responsible for enrollment were different from those assigning participants to groups to ensure the integrity of allocation concealment.

### Rewarming

Intraoperative management followed our institution's standard of care, which included the routine use of fluid warmers, application of forced-air warming blankets, and continuous core temperature monitoring for all patients. Upon arrival at the PACUs, patients who met the preliminary criteria were assessed. Only those with a core temperature lower than 36 °C were enrolled and randomly allocated into 3 groups (n=44 per group): group C (forced-air warming set to 38 °C), group F1 (forced-air warming set to 42 °C), and group F2 (forced-air warming initially set to 42 °C, then reduced to 38 °C once the core temperature reached 36 °C).

All participants received identical warming systems and core temperature monitoring devices at the PACU. Rewarming was carried out using a forced-air warming system (Bair Hugger) with a blanket and a core temperature measurement device (Braun ThermoScan ear thermometer).

The rewarming process was monitored and recorded every 5 minutes during the active warming phase. In groups C and F1, the forced-air warmer was discontinued once the core temperature reached ≥36 °C, at which point patients were covered with a regular blanket and monitored every 10 minutes. In group F2, the setting was reduced to 38 °C upon reaching a

core temperature of ≥36 °C, and patients were similarly monitored every 10 minutes until discharge from the PACU. Rewarming time was calculated as the time taken for the core temperature to rise from baseline to ≥36 °C, measured in minutes.

### Outcome Measures

The primary outcome was the rewarming time, defined as the duration from the initiation of rewarming to the recovery of normothermia (core temperature ≥36 °C). Additionally, the incidence of a decrease in core temperature after achieving normothermia was recorded in each group.

Secondary outcomes included the incidence of adverse events—such as hypotension, hypertension, arrhythmias, nausea or vomiting, pain, and shivering—and patient satisfaction. Patient satisfaction was evaluated using 2 validated instruments: the 5-point Patient Comfort Scale, which measures overall comfort and satisfaction, and the 7-point Thermal Comfort Scale, which assesses subjective thermal sensation ranging from –3 (cold) to +3 (hot), with 0 representing thermal neutrality.

### Data Collection

Preoperative and intraoperative data were collected, including patient demographics, surgical procedure, operative time, anesthetic technique, blood loss, and fluid and blood product administration. Upon PACU admission, core temperature was recorded every 5 minutes during the rewarming phase by trained PACU nurses. To ensure consistency, the same nurse performed all assessments for a given patient using the same device and the same ipsilateral ear. Blinding of these nurses was not feasible because the FAW device displayed temperature settings during operation; consequently, the nurses were aware of group allocation, although patients remained blinded. Once the core temperature reached ≥36 °C, measurements continued every 10 minutes until discharge based on the Modified Aldrete scoring system. Throughout the PACU stay, adverse events were monitored continuously, and patient comfort and thermal comfort scores were assessed by the nurses at the time of discharge.

### Statistical Analysis

Data were analyzed using SPSS software version 27 (IBM Corp.). Continuous variables were expressed as mean (SD) or median (IQR), depending on the distribution, which was assessed using the Shapiro-Wilk test. Categorical variables were presented as counts and percentages.

For comparisons between groups, one-way ANOVA was used for normally distributed continuous variables, while the Kruskal-Wallis test was applied to non-normally distributed data. Post hoc analyses were performed using Tukey honest significant difference test for ANOVA and Dunn test for the Kruskal-Wallis test, as appropriate. Categorical variables were compared using the chi-square test or Fisher exact test, as required.

Monte Carlo simulation was utilized for the Fisher exact test extension in contingency tables larger than 2×2 where cell counts were sparse (expected count <5), ensuring robust *P* value estimation without violating asymptotic assumptions. Relative risks with 95% CI were reported for significant categorical outcomes. For continuous variables, effect sizes were expressed as Cohen *d* to ensure consistency and enhance clinical interpretability. This was an intention-to-treat analysis, and all randomized patients were analyzed in their assigned groups. A *P* value of <.05 was considered statistically significant.

### Ethical Considerations

This study was approved by the Human Research Ethics Committees of Ramathibodi Hospital, Mahidol University (approval number MURA2023/202) and registered at Thaiclinicaltrials.org on March 14, 2023 (approval number TCTR20231012004). Written informed consent was obtained from all participants before enrollment. Participants' privacy and confidentiality were strictly protected, and all data were deidentified before analysis. No financial compensation was provided to participants for their participation in the study.

## Results

### Baseline Characteristics

A total of 165 patients were assessed for eligibility between April 2023 and May 2024. Thirty-three patients were excluded (2 with ASA physical status >III and 31 normothermic at PACU admission). Finally, 132 patients were included and equally divided into 3 groups (Figure 1): group C (rewarming set to 38 °C, n=44), group F1 (rewarming set to 42 °C, n=44), and group F2 (rewarming set to 42 °C until reaching a core temperature of 36 °C, then reduced to 38 °C, n=44). The baseline characteristics, including age, gender, ASA physical status, BMI, and underlying diseases, were comparable across all groups, as detailed in Table 1.

**Figure 1.** Flow diagram of patient enrollment. This diagram shows the screening and allocation process for all patients. ASA: American Society of Anesthesiologists; PACU: postanesthesia care unit.

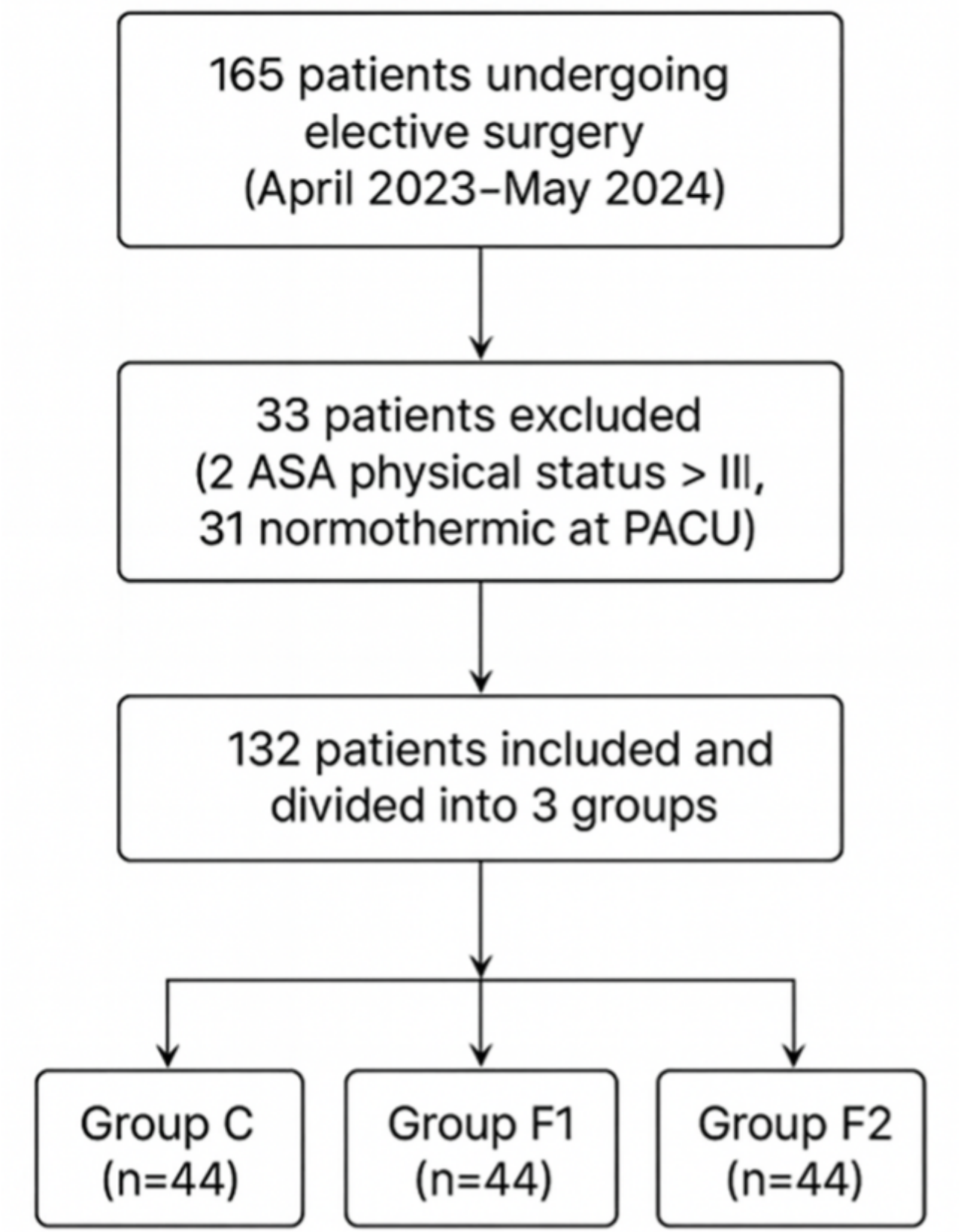


**Table .** Baseline characteristics of patients in the 3 groups.

| Characteristic | Group C (n=44) | Group F1 (n=44) | Group F2 (n=44) | *P* value |
|---|---|---|---|---|
| Age, n (%) | | | | .265 |
| Elderly (age ≥65 y) | 10 (22.73) | 13 (29.55) | 17 (38.64) | |
| Nonelderly (age <65 y) | 34 (77.27) | 31 (70.45) | 27 (61.36) | |
| Gender (male/female) | | | | .083 |
| Male | 17 | 18 | 9 | |
| Female | 27 | 26 | 35 | |
| ASA[a] Physical status, n (%) | | | | .965 |
| >2 | 13 (29.55) | 13 (29.55) | 14 (31.82) | |
| ≤2 | 31 (70.45) | 31 (70.45) | 30 (68.18) | |
| BMI (kg/m$^2$), median (IQR) | 23.85 (22 - 26.95) | 24 (22.43 - 26.3) | 23.67 (21.05 - 27) | .738 |
| Underlying diseases, n (%) | | | | |
| Diabetes mellitus | 3 (6.82) | 8 (18.18) | 7 (15.91) | .259 |
| Hypertension | 14 (31.82) | 18 (40.91) | 19 (43.18) | .511 |
| Obesity | 5 (11.36) | 6 (13.64) | 4 (9.08) | .798 |
| Extreme age | 10 (22.73) | 12 (27.27) | 15 (34.09) | .490 |
| Heart disease | 3 (6.82) | 4 (9.09) | 5 (11.36) | .927 |
| Cerebrovascular disease | 3 (6.82) | 5 (11.36) | 1 (2.27) | .295 |
| Chronic kidney disease | 4 (9.09) | 5 (11.36) | 2 (4.55) | .622 |
| Cancer | 5 (11.36) | 8 (18.18) | 11 (25) | .253 |
| Respiratory disease | 2 (4.55) | 3 (6.82) | 2 (4.55) | >.999 |
| Others | 11 (25) | 14 (31.82) | 7 (15.91) | .217 |

[a]ASA: American Society of Anesthesiologists.

## Intraoperative Data

No significant differences were observed among the 3 groups regarding operative time, type of operation, anesthetic technique, estimated blood loss, total fluid administration, total blood components used, or recorded temperatures, as summarized in Table 2.

**Table .** Intraoperative parameters of patients in the 3 groups.

| Parameter | Group C (n=44) | Group F1 (n=44) | Group F2 (n=44) | *P* value |
|---|---|---|---|---|
| Operative time (min), median (IQR) | 165 (120 - 221.25) | 180 (123.75 - 223.75) | 195 (137.5 - 233.75) | .153 |
| Anesthetic technique, n (%) | | | | .926 |
| Regional | 15 (34.09) | 14 (31.82) | 14 (31.82) | |
| General | 22 (50) | 21 (47.73) | 24 (54.55) | |
| Combined | 7 (15.91) | 9 (20.45) | 6 (13.64) | |
| Operation, n (%) | | | | .731 |
| Open orthopedic surgery | 18 (40.91) | 18 (40.91) | 17 (38.64) | |
| Arthroscopic/laparoscopic orthopedic surgery | 8 (18.18) | 7 (15.91) | 11 (25) | |
| Open gynecologic surgery | 3 (6.82) | 4 (9.09) | 4 (9.09) | |
| Laparoscopic gynecologic surgery | 6 (13.64) | 3 (6.82) | 8 (18.18) | |
| Breast surgery | 3 (6.82) | 0 (0) | 2 (4.55) | |
| Open general surgery | 2 (4.55) | 2 (4.55) | 1 (2.27) | |
| Urological surgery | 1 (2.27) | 1 (2.27) | 0 (0) | |
| Thoracic surgery | 1 (2.27) | 1 (2.27) | 0 (0) | |
| Plastic surgery | 18 (40.91) | 1 (2.27) | 0 (0) | |
| Vascular surgery | 0 (0) | 2 (4.55) | 0 (0) | |
| Laparoscopic general surgery | 2 (4.55) | 4 (9.09) | 1 (2.27) | |
| Otolaryngologic surgery | 0 (0) | 1 (2.27) | 0 (0) | |
| Intraoperative period | | | | |
| Large estimated blood loss (≥500 ml), n (%) | 1 (2.27) | 3 (6.82) | 1 (2.27) | .435 |
| Total fluid administered (ml), median (IQR) | 975 (762.5 - 1425) | 900 (612.5 - 1387.5) | 1000 (712.5 - 1487.5) | .808 |
| Total blood component (ml), median (IQR) | 0 (0) | 0 (0) | 0 (0) | .173 |
| Temperature recorded (°C), mean (SD) | 36.07 (0.43) | 35.93 (0.37) | 36.1 (0.32) | .139 |

## Postoperative Data

Postoperative outcomes recorded in the PACU are summarized in Table 3. No significant differences were observed among the 3 groups regarding tympanic temperature upon arrival or hemodynamic parameters, including blood pressure, heart rate, respiratory rate, and $SpO_2$ ($P$>.05). However, a statistically significant difference was observed in the duration of PACU stay ($P$=.015). Post hoc comparisons revealed significant differences between group C versus group F2 and group F1 versus group F2, indicating a more favorable distribution of discharge times in group F2. Regarding electrocardiogram findings, adverse events were rare, with only 1 patient exhibiting bradycardia.

**Table .** Postoperative outcomes of patients in the 3 groups.

| Parameter | Group C (n=44) | Group F1 (n=44) | Group F2 (n=44) | *P* value |
|---|---|---|---|---|
| Duration in PACU[a] (min),[c] median (IQR) | 60 (60 - 65) | 60 (60 - 60) | 60 (60 - 60) | .015 |
| Tympanic temperature <36 °C upon arrival in the PACU, n (%) | | | | .165 |
| <35 °C | 0 (0) | 1 (2.27) | 3 (6.82) | |
| ≥35 °C | 44 (100) | 43 (97.73) | 41 (93.18) | |
| Systolic blood pressure (mmHg), mean (SD) | 134.07 (22.57) | 134.53 (18.21) | 135.95 (21.86) | .910 |
| Diastolic blood pressure (mmHg), mean (SD) | 76.32 (31.85) | 79.09 (12.5) | 94.45 (106.34) | .341 |
| Respiratory rate (per min), mean (SD) | 17.07 (3.25) | 18.09 (3.2) | 17.95 (3.43) | .290 |
| Heart rate (per min), mean (SD) | 68.61 (12.12) | 72.55 (13.15) | 74.75 (13.31) | .081 |
| $SpO_2$[b] (%), median (IQR) | 100 (99 - 100) | 100 (99 - 100) | 100 (99 - 100) | .688 |

[a]PACU: postanesthesia care unit.

[b]$SpO_2$: peripheral capillary oxygen saturation.

[c]A statistically significant difference was observed only in the duration of stay in PACU (*P*=.015; post hoc comparisons revealed significant differences for group C vs group F2 [Cohen *d*=6.29] and group F1 vs group F2 [Cohen *d*=8.55]).

## Effect of Different Rewarming Methods

Rewarming outcomes across all dimensions are summarized in Table 4. The time to achieve normothermia is illustrated in Figure 2. Figure 3 shows the number of patients who experienced a drop in core temperature below 36 °C after achieving normothermia, and those who required rewarming for more than 1 hour.

No significant differences were observed among the 3 groups in hemodynamic parameters recorded in the PACU, including systolic and diastolic blood pressure, respiratory rate, heart rate, peripheral capillary oxygen saturation, and electrocardiogram findings (*P*>.05).

While the rewarming time did not differ significantly among the groups, the incidence of patients experiencing a drop in core temperature below 36 °C after achieving normothermia was significantly lower in group F2 compared to groups C and F1 (*P*=.009, Table 4). Patients exhibiting temperature decline required extended thermal support to restore or maintain normothermia. Consequently, a significantly higher proportion of patients in groups C and F1 required active warming for more than 1 hour compared to group F2 (*P*=.017). However, this prolonged warming requirement did not lead to a clinically relevant delay in discharge, as the median duration of PACU stay remained 60 minutes across all groups (Table 3).

**Table .** Rewarming outcomes of patients in the 3 groups.

| Parameter | Group C (n=44) | Group F1 (n=44) | Group F2 (n=44) | *P* value |
|---|---|---|---|---|
| Mean rewarming time[a] (min), mean (SD) | 37.39 (16.58) | 35.11 (15.64) | 33.30 (13.81) | .460 |
| Decrease temperature below 36 °C after achieving normothermia, n (%) | 7 (15.91) | 11 (25) | 1 (2.27) | .009 |
| Warming more than 1 h, n (%) | 10 (22.73) | 7 (15.91) | 1 (2.27) | .017 |

[a]For the mean rewarming time, no significant differences were observed among groups; the mean differences (95% CI) compared to group C were −2.28 (−9.03 to 4.47) for group F1 and −4.09 (−10.46 to 2.28) for group F2. However, a statistically significant difference was observed in the proportion of patients with temperature decrease below 36 °C (*P*=.009; significant pairwise differences were observed for group F1 vs group F2 [relative risk (RR)=11.00; 95% CI 1.48 - 81.61]) and those requiring warming for more than 1 h (*P*=.017; significant pairwise differences were observed for group C vs group F2 [RR=10.00; 95% CI 1.34 - 74.84]).

**Figure 2.** Comparison of rewarming outcomes among the 3 study groups. Box plot showing the distribution of time to achieve normothermia. The horizontal line within each box represents the median rewarming time. The top and bottom boundaries of the boxes indicate the IQR, and the whiskers extend to the minimum and maximum values. No statistically significant differences were observed (*P*=.460). Control group (group C): forced-air warming at 38 °C; group F1: forced-air warming at 42 °C; group F2: forced-air warming initially set at 42 °C, then reduced to 38 °C upon reaching 36 °C.

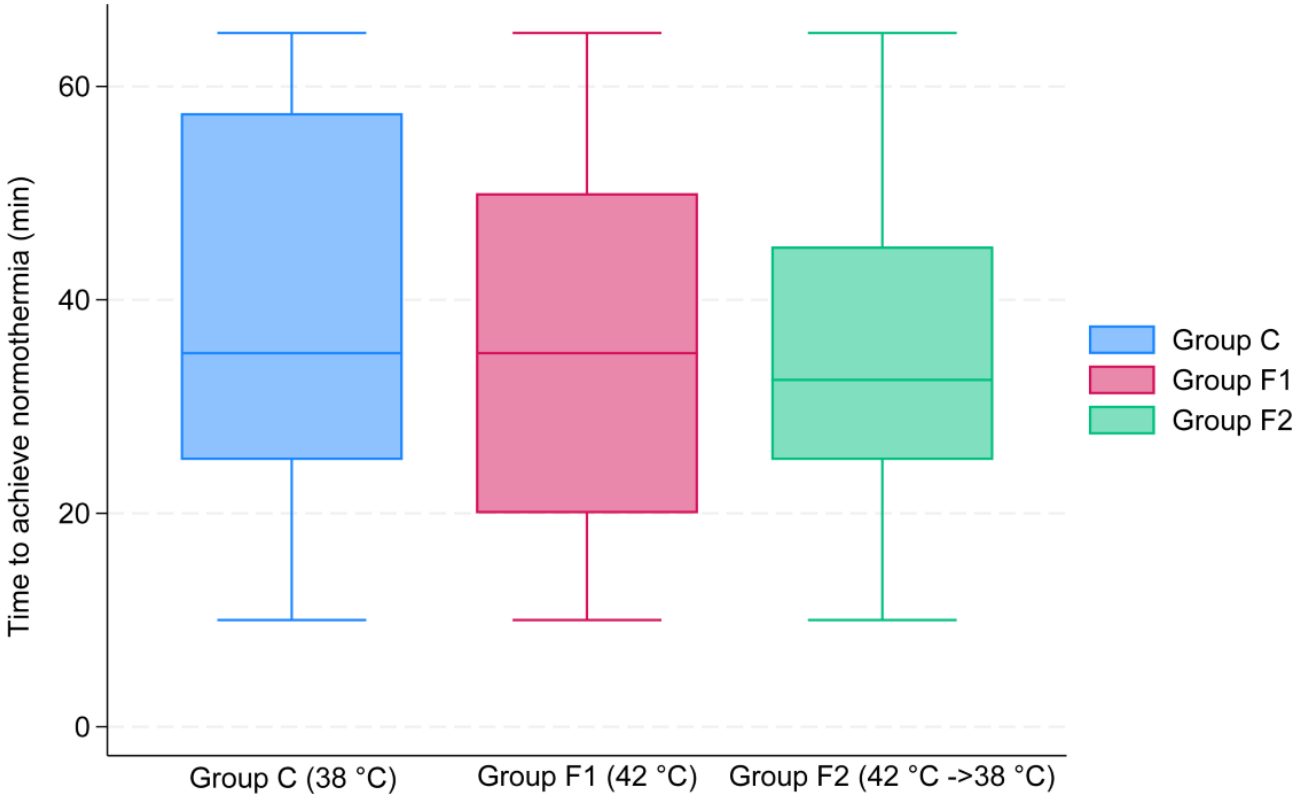
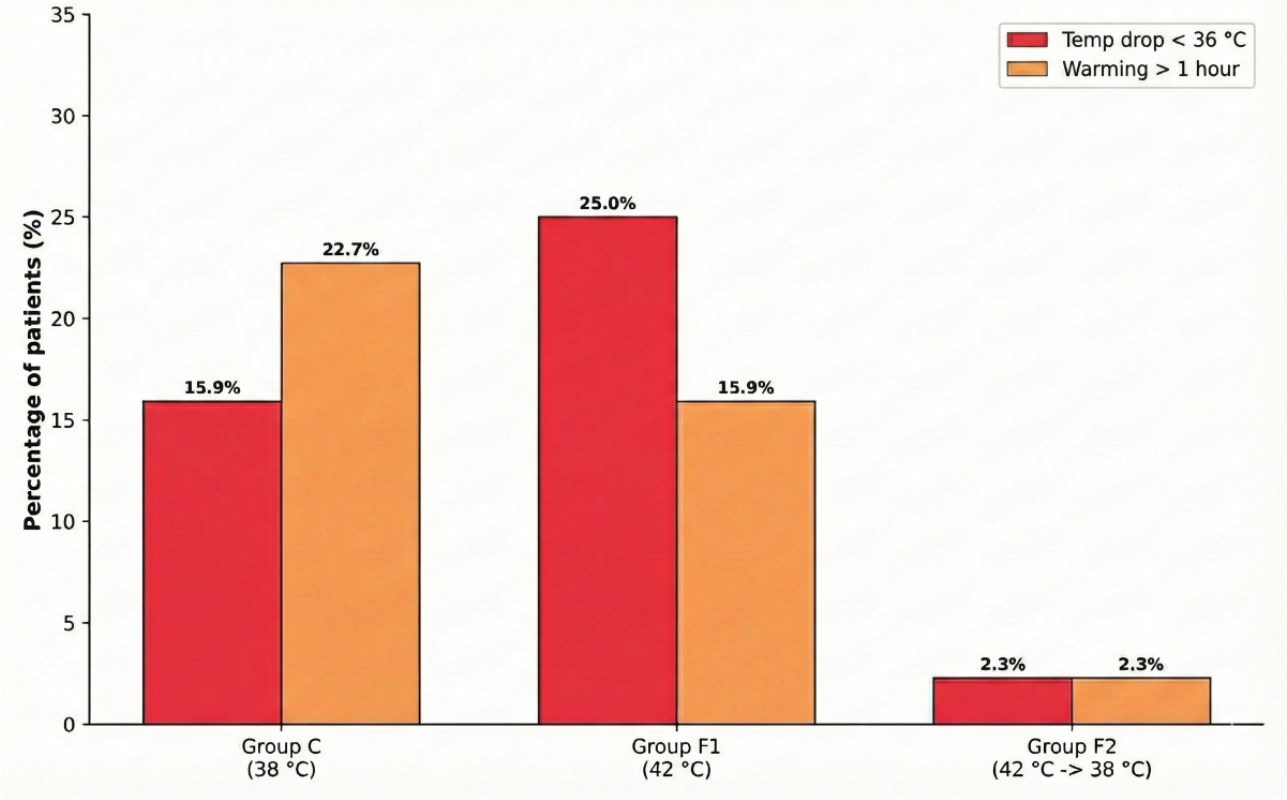


**Figure 3.** Comparison of rewarming outcomes among the 3 study groups. Clustered bar chart illustrating the incidence of recurrent hypothermia (core temperature dropping <36 °C after achieving normothermia) and the proportion of patients requiring active warming for more than 1 h. Group F2 demonstrated significantly lower rates for both outcomes compared to groups C and F1 (*P*=.009 and *P*=.017, respectively). Control group (group C): forced-air warming at 38 °C; group F1: forced-air warming at 42 °C; group F2: forced-air warming initially set at 42 °C, then reduced to 38 °C upon reaching 36 °C.




## Adverse Events

In terms of postoperative adverse events, no significant differences were observed among the 3 groups. Pain was the most frequently reported complication, affecting 22.7% (10/44) of patients in group C and 18.2% (8/44) in both group F1 and group F2 (*P*=.826). Nausea and vomiting occurred infrequently, with an incidence ranging from 2.3%(1/44) to 4.6%(2/44) across the groups (*P*>.999). Shivering was reported in 4.6% (2/44) of patients in group C, 6.8% (3/44) in group F1, and 2.3% (1/44) in group F2 (*P*=.871). Hemodynamic events were rare, comprising 1 case of hypertension in group C, no events in group F1, and 1 case each of hypotension and hypertension in group F2 (all *P*>.999). No patients experienced arrhythmia or other adverse effects. Overall, the incidence of postoperative complications was low and comparable among the groups, supporting the safety of forced-air warming across different temperature settings.

Patient comfort, as measured by satisfaction levels, showed a significant difference among the groups (*P*=.049), along with the average comfort scores (*P*=.039). The proportion of patients reporting being “very much satisfied” was 27.27% (12/44) in group C, 43.18% (19/44) in group F1, and 52.27% (23/44) in group F2 (Table 5). However, there was no significant difference in the thermal comfort scale among the groups (*P*=.131).

**Table .** Patient satisfaction in the 3 groups.

| Parameter | Group C (n=44) | Group F1 (n=44) | Group F2 (n=44) | *P* value |
|---|---|---|---|---|
| Patient's comfort, n (%) | | | | .049 |
| Very much satisfied | 12 (27.27) | 19 (43.18) | 23 (52.27) | |
| Somewhat satisfied | 29 (65.91) | 21 (47.73) | 21 (47.73) | |
| Undecided | 2 (4.55) | 4 (9.09) | 0 (0) | |
| Not really satisfied | 0 (0) | 0 (0) | 0 (0) | |
| Not at all satisfied | 1 (2.27) | 0 (0) | 0 (0) | |
| Patient's comfort (average score), median (IQR) | 4 (4-5) | 4 (4-5) | 5 (5-5) | .039 |
| Thermal comfort scale, n (%) | | | | .131 |
| Hot | 1 (2.27) | 4 (9.09) | 2 (4.55) | |
| Warm | 34 (77.27) | 32 (72.73) | 31 (70.45) | |
| Slightly warm | 6 (13.64) | 2 (4.55) | 9 (20.45) | |
| Neutral | 2 (4.55) | 6 (13.64) | 2 (4.55) | |
| Slightly cold | 1 (2.27) | 0 (0) | 0 (0) | |
| Cool | 0 (0) | 0 (0) | 0 (0) | |
| Cold | 0 (0) | 0 (0) | 0 (0) | |
| Thermal comfort scale (average score), median (IQR) | 2.00(2-2) | 2.00(2-2) | 2.00 (1.25 - 2) | .676 |

## Discussion

### Principal Findings

Postoperative hypothermia is a frequent complication of both general and regional anesthesia, primarily resulting from thermoregulatory impairment and internal heat redistribution [21]. While previous research identified FAW at 42 °C as effective for elderly patients [19], evidence regarding the optimal temperature setting for the general surgical population remains limited. Consequently, this trial aimed to evaluate the most effective and efficient rewarming protocol for patients undergoing various surgical procedures.

In this study, hypothermia was defined as a core temperature <36 °C upon PACU admission. Tympanic thermometry was selected over invasive nasopharyngeal or rectal probes, as used in previous studies [19,22], to prioritize patient comfort during the awake recovery phase.

The principal finding of this study is that increasing the FAW setting from 38 to 42 °C did not yield a statistically significant reduction in the overall rewarming time to normothermia. While group F2 achieved the target temperature approximately 3 to 4 minutes faster than the control group, the precision estimates provided by the 95% CI suggest that this difference is negligible. Given that all groups achieved normothermia within a comparable timeframe, the variation in rewarming speed appears to lack clinical relevance for PACU throughput.

Several physiological factors may explain why higher settings did not produce faster rewarming, a finding that contrasts with some previous studies [19]. Peripheral vasoconstriction can limit the rate of convective heat transfer from the skin to the core, creating a "plateau effect" regardless of the external heat gradient provided by higher FAW settings. Moreover, as the core temperature approaches the normal thermoregulatory threshold, the body initiates vasodilation to redistribute heat, preventing a linear increase in core temperature [21,23]. Device-specific factors, such as automatic safety regulation at higher settings or variability in blanket positioning, may have further minimized the actual difference in heat delivery.

Although rewarming rates were comparable, the group F2 protocol demonstrated superior thermal stability. Unlike group C, which exhibited a significantly higher incidence of prolonged rewarming, the step-down protocol (group F2) effectively minimized the number of "outliers"—patients requiring extended care due to thermal instability. The prolonged rewarming observed in group C is likely due to several physiological and thermodynamic factors. At a lower temperature (38 °C), the gradient between the patient's core temperature and the surrounding warming environment is reduced, leading to a slower rate of heat transfer [24]. Additionally, peripheral vasoconstriction limits blood flow to the skin and extremities, impeding the transport of externally applied heat to the core [25]. Furthermore, the reduced metabolic rate associated with hypothermia decreases endogenous heat generation, collectively contributing to the extended recovery time [21]. Consequently, the requirement for prolonged active warming in groups C and F1 likely contributed to the statistical difference observed in the total duration of PACU stay (*P*=.015). Although the median stay was consistent at 60 minutes across all groups, the distribution of discharge times suggests that while group F2 may not shorten the mandatory minimum recovery

time, it optimizes unit throughput by reducing the incidence of prolonged stays.

### Clinical Implications

Regarding safety and comfort, the incidence of adverse events—including pain, nausea, vomiting, hemodynamic changes, and shivering—did not differ significantly among the groups, and no severe adverse events were observed. These results align with the safety profile reported by Xu et al [19] However, regarding patient experience, group F2 reported higher satisfaction scores related to comfort during rewarming compared to groups C and F1. This suggests that an initial high-temperature setting effectively enhances thermal comfort, while the subsequent reduction prevents the discomfort associated with overheating.

Intraoperative factors—including ambient cooling, fluid administration, and anesthesia-induced thermoregulatory impairment—are known to significantly impact rewarming. In this study, potential confounding was minimized through a standardized intraoperative care protocol that included routine fluid and forced-air warming. Furthermore, randomization successfully balanced these physiological stressors across study arms; as shown in Table 2, there were no significant differences in operative duration, total fluid volume, or anesthetic technique. Consequently, the observed differences in PACU outcomes can be primarily attributed to the specific postoperative warming protocols rather than intraoperative disparities.

### Strengths and Limitations

This study has notable strengths and limitations. A key strength is the rigorous randomization and standardized intraoperative care, which successfully balanced potential confounders such as operative duration and fluid volume across study arms. However, several limitations exist. First, strict environmental control of the PACU was challenging due to the open-plan nature of the unit. The ambient temperature fluctuated between 22 and 24 °C, which serves as a potential environmental confounder influencing convective heat loss. Nevertheless, this variation reflects real-world clinical conditions, potentially enhancing the ecological validity of our results. Second, regarding measurement, reliance on tympanic thermometry introduces inherent variability compared to the gold standard of invasive core monitoring. We acknowledge that readings can be affected by factors such as probe positioning, cerumen obstruction, and post-anesthetic peripheral vasoconstriction. To mitigate these inaccuracies, we strictly standardized the technique by using the same device and assessing the ipsilateral ear throughout the study, aiming to balance measurement precision with patient comfort in the awake state. Third, data collection involved intermittent recordings at 5-minute intervals rather than continuous electronic monitoring. While this frequency is clinically practical, it may lack the temporal resolution to capture rapid, transient temperature fluctuations during the active rewarming phase, potentially masking the true extent of thermal variability. Finally, we did not perform a formal cost-effectiveness analysis. Although the step-down protocol (group F2) showed potential for optimizing PACU throughput, future studies including economic evaluations are needed to confirm the financial implications of these warming strategies.

### Conclusions

In summary, although varying temperature settings of forced-air warming systems produced comparable rewarming times, the protocol involving an initial setting of 42 °C followed by a reduction to 38 °C (group F2) was associated with superior maintenance of normothermia and a significantly lower incidence of postoperative hypothermia recurrence. These findings underscore the potential benefits of implementing optimized warming protocols to enhance patient outcomes in the PACU.

## Acknowledgments

The authors would like to thank all patients who willingly participated in this study, without whom the research would not have been possible. We gratefully acknowledge the contribution of Ms. Rojnarin Komonhirun, M.Ed., for her valuable assistance with data analysis. We also thank the PACU staff of Ramathibodi Hospital for their support in facilitating the data collection process.

## Funding

No external financial support or grants were received from any public, commercial, or not-for-profit entities for the research, authorship, or publication of this article.

## Data Availability

Due to privacy concerns, the datasets generated or analyzed during this study are not publicly available. However, deidentified data may be provided by the corresponding author upon reasonable request.

## Authors' Contributions

Conceptualization: KP (lead), SS (equal), RS (supporting), WW (supporting)
Data curation: KP (lead), SS (equal)
Formal analysis: KP (lead), SS (equal)
Investigation: KP (lead), SS (equal)
Methodology: KP (lead), SS (equal)

Resources: KP (lead), SS (equal), WW (supporting)
Supervision: KP (lead), RS (supporting)
Validation: KP (lead), SS (equal)
Visualization: KP (lead), SS (equal), RS (supporting), WW (supporting)
Writing – original draft: KP (lead), SS (equal)
Writing – review & editing: KP (lead), SS (equal), RS (supporting), WW (supporting)

## Conflicts of Interest

None declared.

## References


1. Shenoy L, Krishna HM, Kalyan N, Prasad KH. A prospective comparative study between prewarming and cowarming to prevent intraoperative hypothermia. J Anaesthesiol Clin Pharmacol 2019;35(2):231-235. [doi: 10.4103/joacp.JOACP_353_17] [Medline: 31303714]
2. Ashoobi MT, Shakiba M, Keshavarzmotamed A, Ashraf A. Prevalence of postoperative hypothermia in the post-anesthesia care unit. Anesth Pain Med 2023 Oct;13(5):e136730. [doi: 10.5812/aapm-136730] [Medline: 38028114]
3. Torossian A, Bräuer A, Höcker J, Bein B, Wulf H, Horn EP. Preventing inadvertent perioperative hypothermia. Dtsch Arztebl Int 2015 Mar 6;112(10):166-172. [doi: 10.3238/arztebl.2015.0166] [Medline: 25837741]
4. Shaw CA, Steelman VM, DeBerg J, Schweizer ML. Effectiveness of active and passive warming for the prevention of inadvertent hypothermia in patients receiving neuraxial anesthesia: a systematic review and meta-analysis of randomized controlled trials. J Clin Anesth 2017 May;38:93-104. [doi: 10.1016/j.jclinane.2017.01.005] [Medline: 28372696]
5. Rauch S, Miller C, Bräuer A, Wallner B, Bock M, Paal P. Perioperative hypothermia - a narrative review. Int J Environ Res Public Health 2021 Aug 19;18(16):8749. [doi: 10.3390/ijerph18168749] [Medline: 34444504]
6. Munday J, Duff J, Wood FM, Sturgess D, Ralph N, Ramis MA. Perioperative hypothermia prevention: development of simple principles and practice recommendations using a multidisciplinary consensus-based approach. BMJ Open 2023 Nov 14;13(11):e077472. [doi: 10.1136/bmjopen-2023-077472] [Medline: 37963694]
7. Grant MC, Engelman DT. Enhanced recovery after surgery: overarching themes of the ERAS® Society Guidelines & Consensus Statements for Adult Specialty Surgery. Perioper Med (Lond) 2025;14(1):120. [doi: 10.1186/s13741-025-00590-0] [Medline: 41168801]
8. Wongsripuemtet P, Leelanokrom R, Tangwiwat S, Vorapaluk P, Karnjanarachata C, Karnjanawanichkul O, et al. Clinical guidelines for prevention and management of perioperative hypothermia by the Royal College of Anesthesiologists of Thailand. Thai J Anesthesiol 2024;50:287-300 [FREE Full text]
9. Lynch S, Dixon J, Leary D. Reducing the risk of unplanned perioperative hypothermia. AORN J 2010 Nov;92(5):553-562. [doi: 10.1016/j.aorn.2010.06.015] [Medline: 21040819]
10. Bräuer A, Quintel M. Forced-air warming: technology, physical background and practical aspects. Curr Opin Anaesthesiol 2009 Dec;22(6):769-774. [doi: 10.1097/ACO.0b013e328331d134] [Medline: 19734783]
11. Warttig S, Alderson P, Campbell G, Smith AF. Interventions for treating inadvertent postoperative hypothermia. Cochrane Database Syst Rev 2014 Nov 20;2014(11):CD009892. [doi: 10.1002/14651858.CD009892.pub2] [Medline: 25411963]
12. Pruitthitada S, Sinchai S. Effective of preoperative force air warming on perioperative hypothermia in patients undergoing caesarean section under spinal anesthesia. Thai J Anesthesiol 2022;48(2):87-98 [FREE Full text]
13. Becerra Á, Valencia L, Villar J, Rodríguez-Pérez A. Short-periods of pre-warming in laparoscopic surgery. A non-randomized clinical trial evaluating current clinical practice. J Clin Med 2021 Mar 3;10(5):1047. [doi: 10.3390/jcm10051047] [Medline: 33802512]
14. Jun JH, Chung MH, Kim EM, et al. Effect of pre-warming on perioperative hypothermia during holmium laser enucleation of the prostate under spinal anesthesia: a prospective randomized controlled trial. BMC Anesthesiol 2018 Dec 22;18(1):201. [doi: 10.1186/s12871-018-0668-4] [Medline: 30579334]
15. Yoo JH, Ok SY, Kim SH, et al. Effects of 10-min of pre-warming on inadvertent perioperative hypothermia in intraoperative warming patients: a randomized controlled trial. Anesth Pain Med (Seoul) 2020 Jul 31;15(3):356-364. [doi: 10.17085/apm.20027] [Medline: 33329836]
16. Parameeanon M, Poomepraman W, Sarachai W. Effect of pre-warming by forced air warmer during total hip and knee arthroplasties for perioperative and postoperative hypothermia. Thai J Anesthesiol 2020;46(2):88-95 [FREE Full text]
17. Zhang R, Chen X, Xiao Y. The effects of a forced-air warming system plus electric blanket for elderly patients undergoing transurethral resection of the prostate: a randomized controlled trial. Medicine (Baltimore) 2018 Nov;97(45):e13119. [doi: 10.1097/MD.0000000000013119] [Medline: 30407328]
18. Tyvold SS. Preventing hypothermia in outpatient plastic surgery by self-warming or forced-air-warming blanket: a randomised controlled trial. Eur J Anaesthesiol 2019 Nov;36(11):843-850. [doi: 10.1097/EJA.0000000000001087] [Medline: 31567576]
19. Xu H, Xu G, Ren C, Liu L, Wei L. Effect of forced-air warming system in prevention of postoperative hypothermia in elderly patients: a prospective controlled trial. Medicine (Baltimore) 2019;98(22):e15895. [doi: 10.1097/MD.0000000000015895]

20. Pisitsak C, Virankabutra T, Deewong K, Pornprasertsuk M. Postoperative hypothermia: incidence and associated factors. Thai J Anesthesiol 2011;37:93-103.
21. Sessler DI. Temperature monitoring and perioperative thermoregulation. Anesthesiology 2008 Aug;109(2):318-338. [doi: 10.1097/ALN.0b013e31817f6d76] [Medline: 18648241]
22. Liu X, Shi Y, Ren C, Li X, Zhang Z. Effect of an electric blanket plus a forced-air warming system for children with postoperative hypothermia: a randomized controlled trial. Medicine (Baltimore) 2017;96(26):e7389. [doi: 10.1097/MD.0000000000007389]
23. Sessler DI. Complications and treatment of mild hypothermia. Anesthesiology 2001 Aug;95(2):531-543. [doi: 10.1097/00000542-200108000-00040] [Medline: 11506130]
24. Taylor EE, Carroll JP, Lovitt MA, et al. Active intravascular rewarming for hypothermia associated with traumatic injury: early experience with a new technique. Proc (Bayl Univ Med Cent) 2008 Apr;21(2):120-126. [doi: 10.1080/08998280.2008.11928375] [Medline: 18382749]
25. Ereth MH, Lennon RL, Sessler DI. Limited heat transfer between thermal compartments during rewarming in vasoconstricted patients. Aviat Space Environ Med 1992 Dec;63(12):1065-1069. [Medline: 1456917]

## Abbreviations

**ASA:** American Society of Anesthesiologists
**FAW:** forced-air warming
**PACU:** postanesthesia care unit

*Edited by J Shiffermiller; submitted 10.Oct.2025; peer-reviewed by P Sawaddiruk; revised version received 10.Dec.2025; accepted 25.Dec.2025; published 28.Jan.2026.*

*Please cite as:*
*Pasutharnchat K, Seangrung R, Sirisophaphong S, Wongkum W*
*Forced-Air Warming Temperature Settings for Treating Postoperative Hypothermia in the Postanesthesia Care Unit: Randomized Controlled Trial*
*JMIR Perioper Med 2026;9:e85045*
*URL: https://periop.jmir.org/2026/1/e85045*
*doi:10.2196/85045*

# Virtual Reality for the Management of Postoperative Pain and Anxiety in Children and Adolescents Undergoing Nuss Repair of Pectus Excavatum: Randomized Controlled Trial

Charlotte M Walter[1], MD; Dillon Froass[2], BS; Nora Bell[3], MD; Lauren Haack[4], MD; Chloe Boehmer[1], MA; Claudia Bruguera Torres[1], MD; Rachel Spivak[1], BA; Max Chou[5], MD; Kristie Geisler[1], BS; Keith O'Conor[6], MD; Sara E Williams[7], PhD; Lili Ding[1], PhD; Christopher D King[1], PhD; Vanessa A Olbrecht[8,9], MBA, MD

[1]Cincinnati Children's Hospital Medical Center, 3333 Burnet Ave, Cincinnati, OH, United States

[2]The Ohio State University Wexner Medical Center, Columbus, OH, United States

[3]Department of Pediatrics, Vanderbilt University, Nashville, TN, United States

[4]Department of Surgery, Wayne State University, Detroit, MI, United States

[5]Department of Anesthesiology, University of Cincinnati Medical Center, Cincinnati, OH, United States

[6]Department of Anesthesiology, Penn State Milton S. Hershey Medical Center, Hershey, PA, United States

[7]Department of Anesthesiology, Perioperative and Pain Medicine, Stanford University, Palo Alto, CA, United States

[8]Department of Anesthesiology, Nemours Children’s Hospital, Wilmington, DE, United States

[9]Sidney Kimmel Medical College, Thomas Jefferson University Hospital, Philadelphia, PA, United States

**Corresponding Author:**
Charlotte M Walter, MD
Cincinnati Children's Hospital Medical Center, 3333 Burnet Ave, Cincinnati, OH, United States

## Abstract

**Background:** Virtual reality (VR) is a novel technology with implications for pain and sensory processing. VR may serve as a novel, scalable method to deliver clinically validated therapy for pain management as an alternative or adjunct to opioids for acute pain. Given that psychological factors and pain perception are both components of postoperative pain, it may also be beneficial to incorporate modalities that decrease anxiety, such as active relaxation and guided meditation with VR. Unfortunately, these therapies are not widely available due to multiple barriers. VR has the potential to deliver pain-reducing, psychologically based therapy to children, thereby enhancing multimodal analgesia and potentially decreasing opioid use. This study investigates the role of VR in reducing pain and anxiety after surgery. Given the substantial risks associated with opioid use, particularly in younger populations, alternative pain management strategies are crucial.

**Objective:** The primary aim of this study was to evaluate the efficacy of VR as a nonpharmacological intervention for managing postoperative pain intensity, pain unpleasantness, anxiety, and opioid use in children and adolescents undergoing Nuss repair of pectus excavatum.

**Methods:** A single-center, prospective, randomized, controlled trial was conducted at a tertiary care children’s hospital and research center. Ninety children and adolescents (8-18 y) undergoing the Nuss procedure were randomized to guided relaxation or mindfulness VR (n=30) and distraction-based gaming VR (n=30), combined to form the VR group (n=60), and a control group using a passive 360° video (n=30). Patients received a 10-minute session on postoperative days 1 and 2. Pain intensity, pain unpleasantness, and anxiety were evaluated before and 0-, 15-, and 30-minute post-session. In-hospital pain scores, anxiety scores, and opioid use were collected.

**Results:** Children and adolescents who participated in VR reported a significantly greater decrease in pain intensity from baseline (0.41, SE 0.23) compared with those in the 360° video group at 30 minutes ($P$=.04) before multiplicity adjustment but not after multiplicity adjustment. There were no significant differences in pain scores or opioid use between the VR and control groups on postoperative day 1 or 2, nor were there changes in pain unpleasantness or anxiety at any time after the intervention.

**Conclusions:** Daily, 10-minute VR sessions provided some trends toward transient analgesic and anxiolytic effects, albeit none that were statistically significant. VR did not significantly decrease overall pain scores or opioid usage, possibly due to the limited intervention duration and high standardized opioid use. Future studies should investigate extended and more frequent VR sessions and the integration of VR with other therapeutic modalities.

**Trial Registration:** ClinicalTrials.gov NCT04351776; https://clinicaltrials.gov/study/NCT04351776

***(JMIR Perioper Med 2026;9:e80902)*** doi:10.2196/80902

**KEYWORDS**

analgesia; anxiety; distraction-based virtual reality; pediatric anesthesia; prospective studies; child; human; virtual reality; pediatric pain medicine; acute postoperative pain

# Introduction

## Background

Multimodal pain management techniques for acute postoperative pain are commonly studied and utilized [1]; opioids continue to be the cornerstone of postoperative pain management. Opioid misuse continues to be a major public health issue in the United States, with children and adolescents particularly vulnerable, as many are initially exposed to opioids prescribed for pain management [2-4]. Furthermore, the risk of future opioid overdose significantly increases with the quantity of pills prescribed; adolescents receiving 30 or more pills have a 35% higher rate of overdose than those prescribed 18 or fewer pills [5]. The prescription of opioid analgesics is a well-documented pathway to misuse, opioid use disorder, and overdose [6].

The Nuss procedure, performed to repair pectus excavatum, is associated with severe postoperative pain [7]. Effective management of postoperative pain after this surgery is crucial, as alleviating pain can enhance patient satisfaction and reduce complication rates [8]. Effective pain management techniques and regimens vary across pediatric institutions and have begun including intercostal nerve cryoablation [8,9]. Opioid use during recovery from the Nuss procedure is common, with one study finding that patients used opioids for a median of 8 days with an IQR of 6 - 10 days [10]. Given the absence of standardized postoperative pain management protocols and the high usage of opioids following the Nuss procedure, it is essential to explore nonpharmacologic pain control adjuncts for these patients.

Virtual reality (VR) technology provides an immersive, multisensory, and 3D environment that enables individuals to experience a modified reality, creating a sense of “presence” for each individual [11]. There is a clear need for alternative pain management methods, including nonpharmacologic techniques. VR has been shown to be effective in reducing perioperative and postoperative anxiety in pediatric patients. Studies show significant reductions in anxiety in pediatric patients immediately after distraction-based gaming virtual reality (VR-D) sessions, with some effects lasting for at least 15 minutes post-intervention [12,13]. Two approaches—VR-D and guided relaxation–based virtual reality (VR-GR)—are being researched for their effectiveness in reducing pain and anxiety following surgery.

VR-D immerses patients in engaging experiences that help divert attention from pain or anxiety, providing effective short-term relief. Integration of these techniques is challenging in the perioperative space, with limited providers and resources and high costs limiting its feasibility. VR can be used anywhere, anytime, with access to a headset.

Gate control theory suggests that distraction can be a valuable tool for pain management, as attentional load is fixed, and distraction toward a pleasant experience means less attention to pain [14,15]. It has been associated with immediate and short-term reductions in postoperative pain intensity and unpleasantness. VR-D techniques have been shown to decrease acute pain in children and adults [16-18]. Single sessions of VR-D have been shown to reduce postoperative pain for up to 30 minutes in some cases, regardless of baseline pain catastrophizing levels, suggesting broad applicability across pediatric populations experiencing postoperative pain [19]. The use of VR-D has also demonstrated pain reduction comparable to opioid use in burn injury patients during wound cleaning [20]. While VR-D is particularly effective for short-term pain management, additional strategies like guided relaxation may be needed for longer-lasting pain relief [21].

VR-GR seeks to provide more sustained pain relief by combining distraction with mind-body techniques, such as guided relaxation or mindfulness within the VR environment. Psychological factors—including calmness, fear, anxiety, and depression—affect the subjective experience of pain [22]. Resilience has been negatively associated with pain unpleasantness, potentially serving as a protective factor in patients with higher baseline anxiety [22,23]. Incorporating active relaxation and guided meditation techniques may significantly contribute to pain reduction. This combination of settling the mind to increase resilience and distraction from acute pain may play a significant role in acute pain reduction [21]. VR-GR may further improve anxiety reduction, especially in children with higher anxiety sensitivity [12,24]. Although VR-GR may offer additional benefits for sustained pain relief compared to distraction alone, its effects were also primarily transient [21]. Using VR to perform guided relaxation could expand the benefits of these nonpharmacological pain management techniques to more children, including those having surgery.

Overall, VR is a promising nonpharmacologic tool for managing postoperative pain and anxiety in children and adolescents. It can potentially enhance the perioperative experience, reduce reliance on pharmacological interventions, and increase patient and family satisfaction. However, randomized controlled trials are needed to establish standardized protocols and explore VR integration with other therapies, such as biofeedback, for more durable outcomes [13,21,25].

## Aim

In this prospective, randomized, controlled clinical trial, we compare the short-term efficacy of immersive VR in decreasing acute postoperative pain (primary outcome), anxiety, and opioid consumption following pectus excavatum repair. We hypothesize that the use of VR will be more effective at reducing pain, anxiety, and opioid use as compared to the control group in this patient population.

# Methods

The original protocol for this study has been published [26].

## Study Design and Setting

This single-center, randomized, unblinded clinical trial was conducted at Cincinnati Children's Hospital Medical Center (CCHMC), a 670-bed tertiary care academic children's hospital. The recruitment began on July 10, 2020, and the study was completed on July 30, 2023. The COVID-19 pandemic delayed study completion. We recruited children and adolescents undergoing Nuss repair of pectus excavatum to investigate the role of VR in the management of postoperative pain and anxiety.

## Ethical Considerations

This study complies with the Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT) statement [27] and the Consolidated Standard of Reporting Trials (CONSORT) statement [28]. The CCHMC Institutional Review Board approved this study (IRB 2019 - 1090) on November 26, 2019, and it was conducted per the rules and regulations for ethical research. This study was registered at ClinicalTrials.gov on April 3, 2020 (NCT04351776). Written informed parental consent and patient assent (children >11 years) were obtained from all participants before enrollment into this study. Patients received a small stipend for participation. All identifying patient information was kept private and confidential.

## Patients and Recruitment

### *Patients*

This study recruited 90 patients (30 patients per group), ages 8 to 18 years, undergoing Nuss repair of pectus excavatum surgery. Informed parental consent and patient assent were obtained before enrollment into this study.

The inclusion criteria were as follows: patients were (1) between the ages of 8 and 18 years, (2) able to read, understand, and speak English, (3) presenting for Nuss repair of pectus excavatum, and (4) followed by the acute pain service following surgery.

The exclusion criteria were as follows: patients with (1) a history of developmental delay, uncontrolled psychiatric conditions, or neurological conditions, (2) a history of seizures, epilepsy, vertigo, or significant motion sickness/nausea/vomiting, or (3) any condition that would preclude the application of the VR headset, such as craniofacial abnormalities.

### *Recruitment*

Approximately 150 Nuss repair surgeries are performed at CCHMC each year. Therefore, our recruitment target of 90 patients was well within achievable limits. During the study, patients who underwent Nuss repair were recruited continuously until we met the targeted enrollment. The operating room schedule and surgical patient list were reviewed for potentially eligible patients, who were approached for recruitment before surgery. If patients wished to participate, consent (and assent for patients >11 years of age) was obtained, and eligibility criteria were verified. We recruited about 2 patients per week. Recruiting stopped during the COVID-19 pandemic, when elective surgeries were not performed, delaying study completion.

## Randomization

Potential patients were identified using the operating room schedule and the pectus surgery list provided by the surgery team. Eligible participants were randomized (1:1:1) into three groups: active distraction-based guided relaxation virtual reality (VR-DGR, n=30) and active VR-D (n=30)—collectively the VR group (n=60)—and a control group—passive 360° video (360-V) without instructions, sound, guided relaxation, or active patient involvement (n=30).

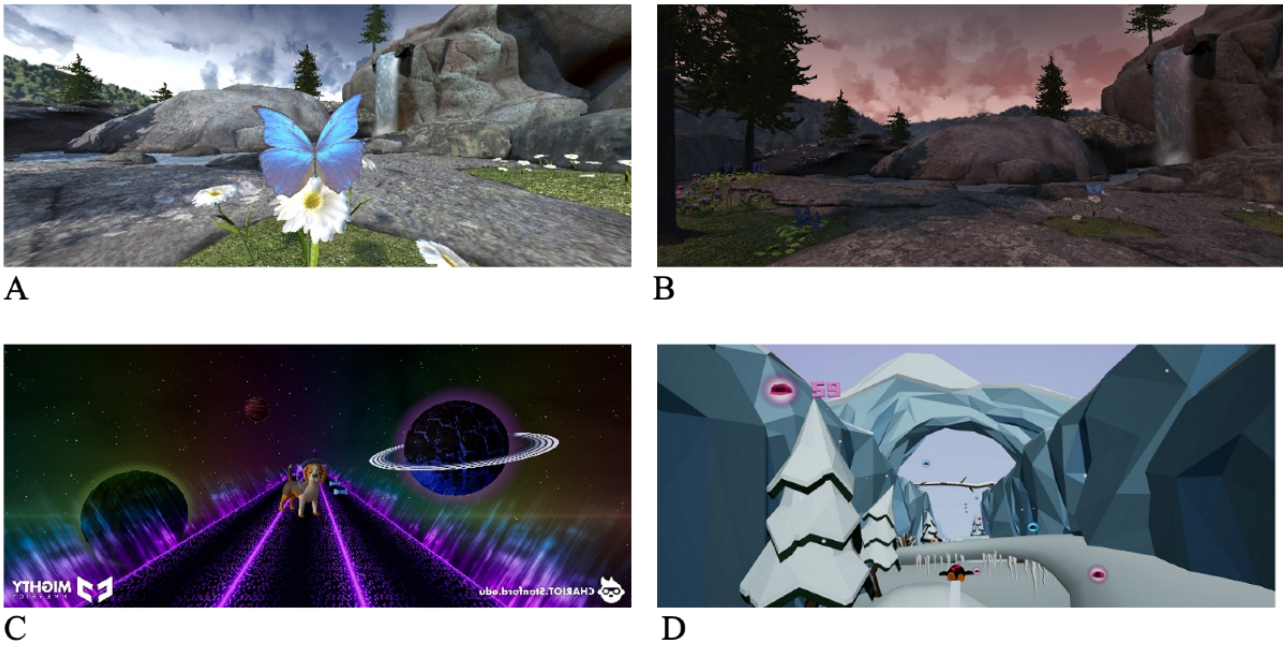
## VR Technology

All participants used a Starlight Xperience VR all-in-one device and software developed specifically for hospital settings. This technology is a customized version of the Lenovo Mirage Solo with a Daydream VR headset. It is easy to disinfect to comply with hospital infection safety protocols. Importantly, an integrated headphone device provides audio content, and the patients use head movements and a handheld controller for interaction and navigation. It is commercially available (not Food and Drug Administration–regulated) and was supplied by the Starlight Children's Foundation.

VR-DGR and 360-V participants used the Mindful Aurora application, developed by the Stanford University Childhood Anxiety Reduction through Innovation and Technology program, to deliver relaxation/mindfulness content, which presents a relaxing nature scene with prompts instructing patients to actively slow and pace their breathing in conjunction with the movement of objects in the VR environment. 360°-V participants experienced the same relaxing nature scene without guided relaxation prompts; the 360°-V group also did not receive any audio and thus did not experience an immersive environment.

VR-D participants had the option to choose and play one of the three games: Space Pups, in which the participant controls a puppy in space and collects treats to music; Pebbles the Penguin, in which the participant controls a penguin sliding down a mountain to collect pebbles; or Wonderglade, in which the participant can play five different mini-carnival games (Figure 1).

**Figure 1.** Scenes from the Mindful Aurora application used in distraction-based guided relaxation virtual reality (VR-DGR) and 360° video (360-V) (A and B), and scenes from Space Pups (C) and Pebbles the Penguin (D) used in distraction-based gaming virtual reality (VR-D).



## Procedure

Consent and assent were obtained before the visit. Patient characteristics, demographics, weight, and pain scores were collected preoperatively. All patients enrolled in this study received standard postoperative care via the CCHMC Pectus Surgery Pain Management Protocol, which standardizes all medications received by all pectus patients. This includes non-opioid pain medication such as pregabalin, acetaminophen, ketorolac, methocarbamol, and diazepam. All participants received the same non-opioid medications. Before the first session, patients completed the Childhood Anxiety Sensitivity Index to establish baseline anxiety levels and the Pain Catastrophizing Scale (PCS) for children. Patients were visited daily for one 10-minute session. Every effort was made to ensure the consistent timing of the visits for all patients. Sessions were completed beginning on postoperative day (POD) 1, then daily until the day of discharge, or until POD 3.

Patients were trained to use the technology before the first VR session. All participants received a device tutorial that taught them how to use the device and introduced them to the VR software. Patients received a script about VR-GR, VR-D, or 360-V, depending on the group to which they are assigned. During each session, patients completed a 10-minute session of either VR-GR, VR-D, or 360-V, per assigned group. Patients were asked to rate their pain intensity, pain unpleasantness, and anxiety via Numerical Rating Scale (NRS), before, immediately after, and 15 and 30 minutes after each session. Pain and anxiety scores and opioid use/day were recorded in REDCap (Research Electronic Data Capture).

## Data Collection

The primary outcome measure was pain intensity (NRS), measured before, immediately after, and 15 and 30 minutes following each session on POD1 and POD2. Secondary outcomes included opioid use on POD0, POD1, and POD2 and pain area under the curve (AUC) on POD1 and POD2, pain unpleasantness, and anxiety scores before and 0, 15, and 30 minutes after each session on POD1 and POD2 to establish change from baseline.

For each eligible participant, data were collected from their patient history/interview and the electronic medical record in a standardized case report form in the REDCap system. Inpatient opioid use was identified from the patient's electronic medical record based on documentation in the medication administration record and transferred to REDCap. All opioid quantities were translated to morphine milligram equivalents (MME) and summed to determine total morphine equivalents per 24-hour period (midnight-to-midnight) during the patients' inpatient stay. Measures used in the study are summarized in Table 1.

**Table .** Scales and questionnaires used in the study.

| Scales and questionnaires | Definition |
|---|---|
| Pain intensity and pain unpleasantness | |
| Numerical Rating Scale (NRS) [29] | The NRS is the most common validated self-report measure of pain intensity and pain unpleasantness. It involves verbally asking for an estimate of pain intensity using numbers from 0 (no pain) to 10 (maximal pain). Pain was described as being like listening to music; pain intensity is the volume, and pain unpleasantness is how much the music is disliked [30]. It requires no equipment to administer or score. |
| Pain intensity across all postoperative days | Area under the time-pain score (NRS) curve using the trapezoidal rule (pain AUC[a]) measured pain intensity across all postoperative days 1 and 2. |
| Anxiety | |
| Pain Catastrophizing Scale for Children (PCS-C) [31] | A validated 13-item questionnaire (each rated on a 5-point scale, 0-4) designed to measure pain catastrophizing in children of age 8 - 17 years. It is adapted from the adult version and assesses three key aspects of pain-related negative thinking: rumination, magnification, and helplessness. |
| Child Anxiety Sensitivity Index (CASI) [32] | A validated 18-item survey that measures perceived anxiety symptoms. Participants respond to each item on a 3-point scale (eg, “none,” “some,” and “a lot”). The total score is calculated by summing the responses, with higher scores indicating greater anxiety sensitivity. The total scores range from 18 to 54. CASI has been used in VR[b] studies in adolescents of age 10 - 21 years [33]. |
| Numerical Rating Scale-Anxiety (NRS-A) [34] | A validated self-report numeric 0 - 10 anxiety scale that is easy to administer to children. The NRS-A is easy to administer and can be used quickly to assess anxiety levels. |
| Opioid use | |
| NIH morphine milligram equivalents (MME) per day [35,36] | Standardizes a metric for quantifying and comparing doses of different opioids. However, MMEs serve as a common metric for comparing different opioids. |

[a]AUC: area under the curve.

[b]VR: virtual reality.

## Statistical Analysis

### *Sample Size Calculation*

Sample size was based on the feasibility of conducting this clinical study and unpublished preliminary data that assessed the impact of a single VR-D session on pain intensity in children and adolescents after surgery, with a goal of 80% power to detect differences in mean changes of 1 between VR and 360-V (given pilot data which showed average change in pain intensity of −1 [SD 1.2] and correlation of 0.88). Assuming similar results in the passive control group, a sample size of 30 per group will have 80% power to detect differences in mean changes of 1 between VR-GR and the two control groups. Significance ($\alpha$) is .025 to control for 2 comparisons.

The VR-DGR and VR-D groups were combined into a single VR group for data analysis because both groups utilized active, distraction-based, immersive VR experiences. VR-DGR did not provide participants with feedback on their respiratory or heart rates. Consequently, it functioned as a distraction-based technique and did not significantly differ from the VR-D experience. Therefore, we combined the two groups, as both were fundamentally distraction-based.

A sample size of 60 for the treatment group and 30 for the control group will have 80% power to detect differences in mean changes of 1 between VR and control.

### *Data Analysis*

All statistical analyses were performed using SAS 9.4 (SAS Institute). Patient demographics were described using mean (SD) or median (IQR) for continuous variables, depending on data distribution, and frequency (percentage) for categorical variables and compared between groups using *t* tests, Wilcoxon rank-sum tests, chi-square tests, or Fisher exact tests, as appropriate. Pain AUCs on POD1 and POD2 were calculated as the area under the time-pain score curve using the trapezoidal rule. MMEs, a metric for quantifying and comparing doses of different opioids, were derived for POD0, POD1, and POD2. Change from baseline on pain intensity, pain unpleasantness, and anxiety immediately after and 15 and 30 minutes following each session was calculated as the postinterval value minus the baseline (before session) value on POD 1 and POD2. Mixed effects models for repeated measures were used for pain AUC, MME, and change from baseline on pain intensity, pain unpleasantness, and anxiety outcomes. All mixed effects models included the intervention group and POD as fixed effects and the participant as a random effect. Models for the change from baseline outcomes also included baseline value, time (0-, 15-,

and 30-min post-intervention), and group and time interaction as fixed effects. Missing data in the outcomes were examined for pattern and assumed missing at random and handled using full information maximum likelihood (FIML) for mixed effects models. Sidak adjustment for multiplicity was used for change from baseline in pain intensity between intervention groups at 3 time points (immediately after and 15 and 30 minutes following each session).

## Results

### Participants

Ninety patients were enrolled in the study (60 in VR and 30 in 360-V; Figure 2 ). The participants comprised 73 male and 17 female patients and had an American Society of Anesthesiologists (ASA) Physical Status Classification System score of 1 - 3, with a mean age of 15.5 (SD 1.4) years. The 2 groups had no differences demographically except for a difference in PCS scores (VR: median 18, IQR 15 - 23; VR-360: median 24, IQR 16 - 28; *P*=.04). Patients were primarily male, adolescent, and Caucasian. This is consistent with the demographics of patients with pectus excavatum [37], and these are the patients most likely to undergo the Nuss procedure [38] (Table 2). All patients had at least 1 observation on all repeated-measured outcomes (pain AUC, MME, pain intensity, pain unpleasantness, and anxiety), and all available data were included in the mixed effects models for the outcomes.

**Figure 2.** CONSORT (Consolidated Standards of Reporting Trials) 2010 flow diagram. VR: virtual reality.

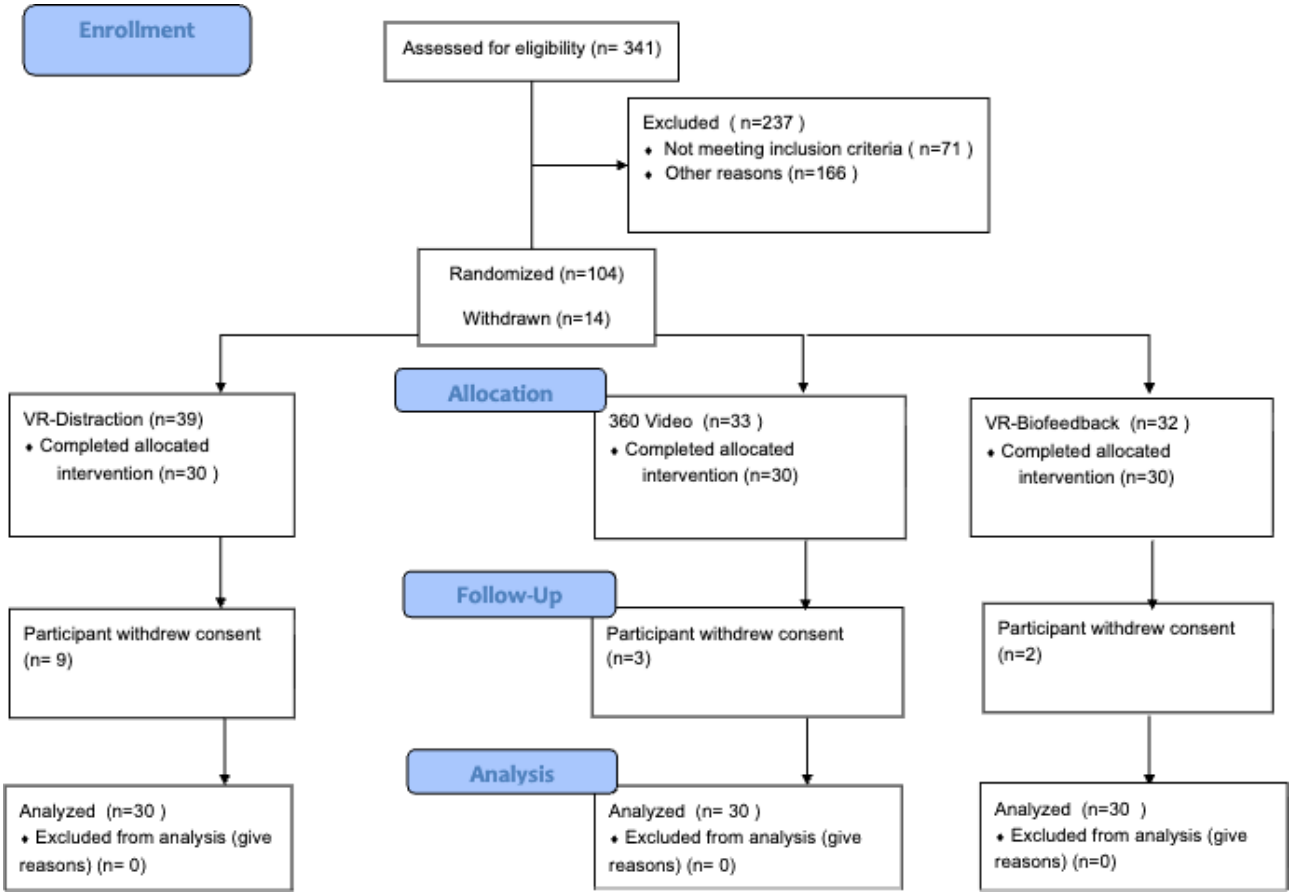


**Table .** Patient characteristics.

| Characteristic | VR[a] | 360° video | Overall | *P* value (test) |
|---|---|---|---|---|
| Age (y), mean (SD) | 15.6 (1.4) | 15.1 (1.5) | 15.5 (1.4) | .10 |
| ASA[b] physical status, n (%) | | | | .34 |
| I | 3 (5) | 4 (13.3) | 7 (7.8) | |
| II | 44 (73.3) | 21 (70.0) | 65 (72.2) | |
| III | 13 (21.7) | 5 (16.7) | 18 (20.0) | |
| Race, n (%) | | | | .55 |
| Caucasian | 57 (95) | 30 (100) | 87 (96.7) | |
| African American | 0 (0) | 0 (0) | 0 (0) | |
| Asian | 0 (0) | 0 (0) | 0 (0) | |
| Other | 3 (5) | 0 (0) | 3 (3.3) | |
| Ethnicity, n (%) | | | | .25 |
| Hispanic | 3 (5) | 0 (0) | 3 (3.3) | |
| Non-Hispanic | 56 (93.3) | 28 (93.3) | 94 (93.3) | |
| Unknown | 1 (1.7) | 2 (6.7) | 3 (3.3) | |
| Sex, n (%) | | | | .70 |
| Male | 48 (80) | 25 (83.3) | 73 (81.1) | |
| Female | 12 (20) | 5 (16.7) | 17 (18.9) | |
| Weight (kg), mean (SD) | 58.9 (9.8) | 56.3 (10) | 58.1 (9.9) | .24 |
| CASI[c] score, mean (SD) | 29.3 (5) | 30.9 (4.8) | 29.8 (5) | .14 |
| Pain Catastrophizing Scale (PCS), median (IQR) | 18 (15-23) | 24 (16-28) | 19 (15-26) | .04 |

[a]VR: virtual reality.

[b]ASA: American Society of Anesthesiologists.

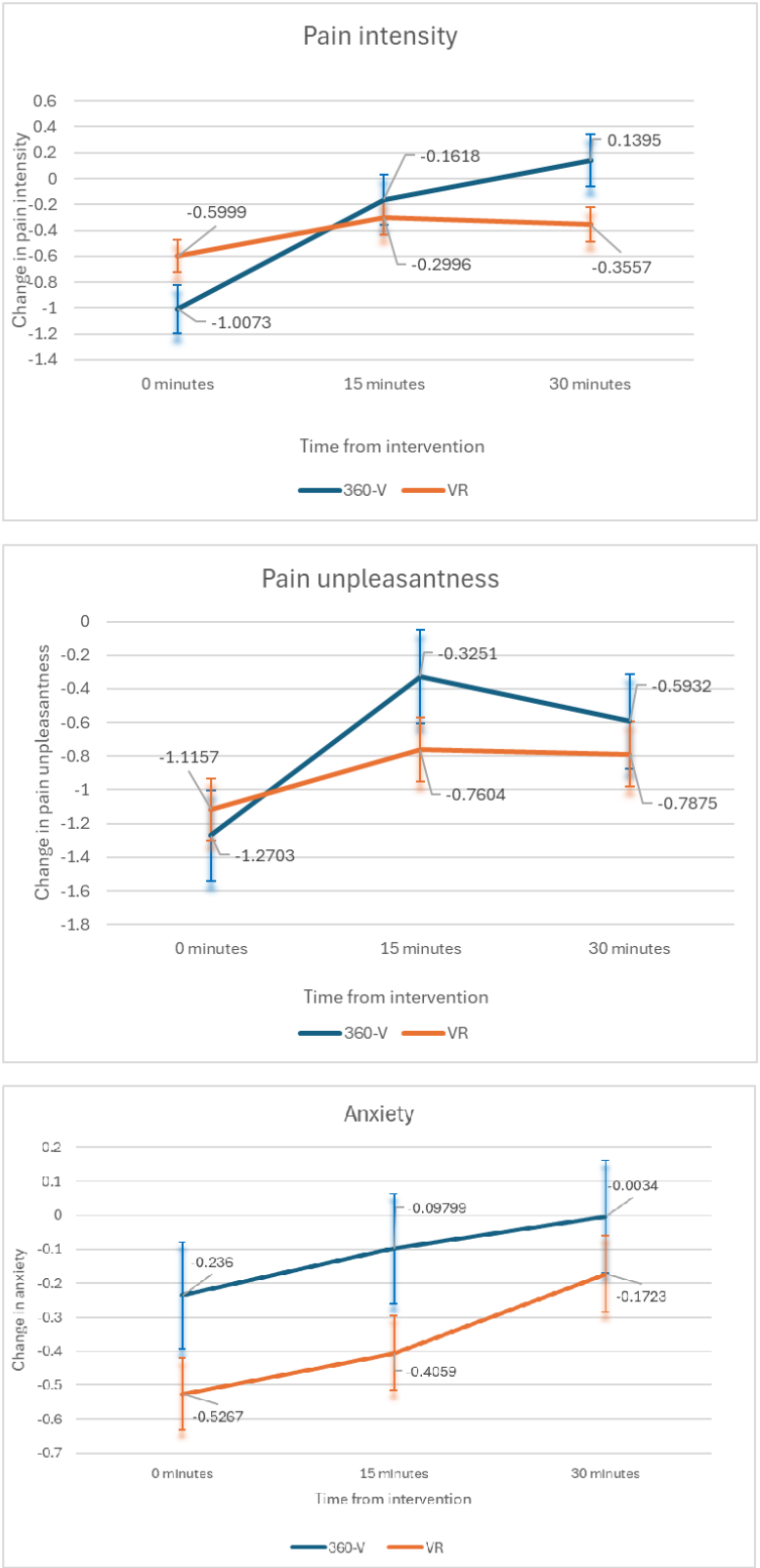
[c]CASI: Child Anxiety Sensitivity Index.

## Changes From Baseline (VR vs Control)

### *Pain Intensity*

Patients who participated in VR reported significantly decreased pain intensity from baseline (0.41 more decrease in pain intensity with SE 0.23) compared with those in the 360-V group at 30 minutes (*P*=.04) before multiplicity adjustment but not after multiplicity adjustment. There was no significant difference from baseline in reported pain intensity between VR vs 360-V immediately following the session (*P*=.08) or after 15 minutes (*P*=.56; Figure 3).

**Figure 3.** Changes in baseline in pain intensity, pain unpleasantness, and anxiety in time points following 360° video (360-V) and virtual reality (VR) in a mixed effect model with standard error bars.




### *Pain Unpleasantness*

There was no significant difference in the reported pain unpleasantness between patients who participated in VR versus 360-V immediately following the VR session ($P$=.64), after 15 minutes ($P$=.20), or after 30 minutes ($P$=.57; Figure 3).

### *Anxiety*

There were no significant differences in reported anxiety from baseline between patients who participated in the VR versus 360-V immediately following the session (*P*=.13), after 15 minutes (*P*=.12), or after 30 minutes (*P*=.40; Figure 3).

### *Inpatient Pain and Opioid Use*

There were no significant differences in mean AUC pain scores between VR and 360-V (*P*=.60). There were also no significant differences in inpatient opioid use (MME/kg/day) between VR and 360-V (*P*=.26; Table 3).

**Table .** Inpatient pain and opioid use.

| Characteristic | VR[a], mean (SD) | 360-V[b], mean (SD) | *P* value[c] |
|---|---|---|---|
| Inpatient pain (AUC[d]) | | | .60 |
| POD[e]1 | 110.0 (34.3) | 107.9 (31.8) | |
| POD2 | 95.9 (32.5) | 90.6 (36.0) | |
| Inpatient opioid use (MME[f]/kg/day) | | | .26 |
| POD0 | 0.18 (0.29) | 0.18 (0.24) | |
| POD1 | 0.58 (0.22) | 0.64 (0.28) | |
| POD2 | 0.50 (0.17) | 0.51 (0.20) | |

[a]VR: virtual reality.

[b]360-V: 360° video.

[c]*P* value from mixed effects models.

[d]AUC: area under the curve.

[e]POD: postoperative day.

[f]MME: morphine milligram equivalents.

## Discussion

### Principal Results

In our study, we found that active, immersive VR experiences had some trends to transient effects on both acute pain and anxiety compared to a nonimmersive 360-V control; however, these effects did not meet statistical significance. Patients who participated in VR reported a significantly greater decrease in pain intensity from baseline (0.41 with SE 0.23) compared with those in the 360-V group at 30 minutes (*P*=.04) before multiplicity adjustment but not after multiplicity adjustment. The trends in reduction in pain and anxiety were small; these trends did not achieve clinical significance either. Current literature indicates that a reduction of at least 2 points on the NRS for pain intensity or a 30% reduction in pain is considered clinically significant [39]. We did not see effects on overall AUC pain scores or opioid use. In this research, the VR-GR experience was likely distraction-based, as we could not document or assess feedback on patients' ability to perform the guided relaxation techniques correctly. In spite of not reaching statistical significance, these trends are not an absence of evidence of the effectiveness of VR to reduce postoperative pain and anxiety. We had relatively similar treatment conditions in small samples with attrition. The trends point us in the direction of future work.

Demographically, our two groups showed no significant differences except for a difference in PCS scores (VR: median 18, IQR 15 - 23; VR-360: median 24, IQR 16 - 28; *P*=.04). However, this result, while statistically significant, may not be clinically significant, as only PCS scores above 30 are clinically relevant, and neither group's median score exceeded 30 [40].

Although we noted some trends toward reduction in acute pain and anxiety from immersive VR following Nuss repair of pectus excavatum, these effects did not result in a significant change in AUC pain scores or inpatient opioid usage between the two groups. Several factors may account for the lack of significance. At our institution, the standard postoperative pain management protocol for following pectus surgery involves the scheduled administration of opioids, meaning all patients receive a standardized, weight-based dosage of opioids during their hospital stay, regardless of their actual pain level, with variations only in as-needed doses [10]. Consequently, opioid consumption may not accurately represent the patients' pain levels and opioid requirements. Home medication use might better reflect patients' pain and opioid needs. This study limited VR experiences to hospitalized patients. Extending its use past hospital discharge may have yielded different results. Future studies should investigate the integration of VR therapy into postoperative pain management both during and after hospitalization.

### Limitations

Although this study was a prospective, randomized clinical trial, which can provide the best clinical evidence and support for VR, it has several limitations due to both study design and factors outside the control of the research team. Our patient population was somewhat homogeneous, as most individuals undergoing pectus excavatum repair are adolescent white males [38]. Therefore, our findings may not be generalizable to a broader population. Additionally, our control group may have been too similar to our intervention group. While the 360-V group did not receive audio instructions for guided relaxation, they still used a VR headset and experienced some level of distraction and immersion. Hence, the difference between the

two groups may not have been substantial enough to detect a meaningful difference. This lack of significant differences in the treatment groups may account for the lack of statistical differences between the two groups when comparing pain and anxiety. Future research will use nonimmersive, non-headset control and retrospectively compare historical data to better assess the impact of VR on these outcomes.

Reporting bias may have also played a role, as study participants may have felt inclined to report decreased pain and/or anxiety after treatment due to their perception of receiving an intervention, regardless of actual changes. The self-reported 1 to 10 rating scale is limited and has been shown to have a moderate correlation with clinical indicators of pain; thus, we cannot rely exclusively on this measure for pain evaluation [41]. Future studies should consider additional outcome measures.

The limited use of the intervention (10 min per d) makes it unlikely to produce meaningful changes in the severe postoperative pain often experienced by these patients. This suggests a need for further investigation of VR for pain relief, potentially incorporating VR interventions more systematically throughout the perioperative period. This could include preoperative VR exposure, repeated interventions multiple times per day, for additional consecutive days following surgery. We accounted for our attrition in our statistical analysis. However, the attrition with small groups to start likely was one reason for our inability to find statistically significant differences in results between the two groups.

## Comparison With Prior Work

Controlling pain after surgery is important; uncontrolled postoperative pain can lead to increased morbidity, decreased function, prolonged recovery, and higher costs [42]. Severe acute postoperative pain can also lead to chronic postoperative pain, with rates of chronic postsurgical pain reported to be about 20% in pediatric populations [43]. Furthermore, opioid use after surgery also has risks, including persistent opioid use postoperatively; one study found a rate of 4.8% of persistent opioid use in postoperative adolescents as compared to a 0.1% rate of persistent opioid use in their nonsurgical matched cohorts [44].

Our study's results align with prior research suggesting that VR-D may play a role in transiently reducing acute pain and that relaxation via guided imagery can promote reductions in both pain and anxiety [20,45]. Research consistently shows that VR is effective in lowering procedural pain, anxiety, and fear in pediatric patients, particularly during needle-related and other painful interventions [46-55]. A few studies have also demonstrated the feasibility of using VR to alleviate acute postoperative pain [21,56]. In adults, a meta-analysis found that patients receiving perioperative VR had lower pain scores than those receiving usual care (mean deviation −0.64, 95% CI −1.05 to −0.22; $P<.02$). Additionally, patients receiving VR postoperatively experienced a significant reduction in pain scores (mean deviation −0.50, 95% CI −0.76 to −0.24; $P=.002$) [57]. One pediatric study indicated that a single preoperative VR experience reduced the need for rescue analgesics in the recovery unit for painful procedures [58].

Relaxation-guided imagery has been shown to reduce both pain and anxiety in children undergoing minor surgery [59]. VR-GR has also demonstrated effectiveness in reducing pain and anxiety in children during medical procedures. These effects were immediate but transient, with some studies reporting reductions lasting up to 30 minutes after a session [21,60,61]. Additionally, a small (n=51) single-center, prospective study evaluated a single VR-GR session for acute postoperative pain and anxiety in children and adolescents. This study showed similar transient reductions in pain intensity and anxiety [21].

Using guided relaxation, our study aimed to harness the benefits of using mind-body techniques to manage postoperative pain. However, there are several potential reasons for the lack of clinically significant lasting effects on acute pain and anxiety in this population or opioid consumption. First, our study did not include a non-headset control group. Previous studies assessing the effects of VR on opioid use compared VR experiences to standard care without VR [55,62,63]. In our study, every patient utilized a VR headset and had some level of an immersive experience, even with 360-V (a nonimmersive option). As a result, all patients, including the control group, likely experienced some distraction.

Moreover, the effectiveness of the VR sessions may have been limited by the short duration of each session and the low number of sessions. Ten minutes per day may be an inadequate time to achieve lasting pain relief of severe acute pain, especially as compared to other therapeutic and/or pharmacologic interventions that are administered more frequently. Future studies should investigate the optimal timing and length of intervention for maximal benefit of VR.

While our results indicated that active, immersive VR experiences had some trends toward effects on both acute pain and anxiety compared to the nonimmersive 360-V control, these effects were not clinically significant. The study may not have produced clear positive results due to experimental factors such as the similarity of the VR and control treatments, the short duration of treatments, and the protocol-driven high use of opioids that were not adequately considered in the study design. Nevertheless, the findings provide a valuable framework for designing future VR studies. We can learn from our null results to design future VR studies with a control treatment that does not use a VR headset and use a population that only receives opioids on an as-needed basis. Longer treatment duration and less subject attrition could also lead to more significant results.

## Conclusions

This study found that daily, 10-minute VR sessions had trends toward transiently reducing pain and anxiety compared to a 360-V experience in participants following Nuss repair of pectus excavatum. These results were not clinically significant. Due to the limited duration of the intervention and the standardized, scheduled, high utilization of opioids in this population, VR was not sufficient in significantly decreasing opioid use and overall AUC pain scores. Despite these conclusions, exploring guided relaxation VR as an adjunct to, rather than a replacement for, postoperative pharmacologic analgesics may prove valuable. Increasing the length and frequency of VR experiences per day, along with a policy of not automatically administering opioids

unless requested, may help decrease opioid usage and AUC pain scores. A systematic integration of VR into perioperative care is likely necessary to impact the pain trajectory and opioid usage in postoperative patients. Furthermore, improving the VR experience to incorporate true guided relaxation would likely enhance effectiveness compared to a purely distraction-based approach. Future studies are needed to further explore the use of this therapy in postoperative pain management.

## Acknowledgments

The authors gratefully acknowledge Maria E Ashton, MS, RPh, MBA, medical writer, Department of Anesthesiology, Cincinnati Children's Hospital Medical Center, for providing writing assistance, editing, and proofreading.

The authors gratefully acknowledge Luke Wilson and the Stanford CHARIOT program for the development and use of the Mindful Aurora application.

## Funding

Support was provided by the Department of Anesthesiology at Cincinnati Children's Hospital Medical Center and a Research Innovation/Pilot Award from Cincinnati Children's Hospital Medical Center.

## Conflicts of Interest

None declared.

Checklist 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File, 1122 KB - periop_v9i1e80902_app1.pdf ]

## References

1. Yaster M, McNaull PP, Davis PJ. The opioid epidemic in pediatrics: a 2020 update. Curr Opin Anaesthesiol 2020 Jun;33(3):327-334. [doi: 10.1097/ACO.0000000000000865] [Medline: 32371640]
2. Drug overdose deaths: facts and figures. National Institute on Drug Abuse (NIDA). URL: https://www.drugabuse.gov/related-topics/trends-statistics/overdose-death-rates [accessed 2026-02-16]
3. Hudgins JD, Porter JJ, Monuteaux MC, Bourgeois FT. Trends in opioid prescribing for adolescents and young adults in ambulatory care settings. Pediatrics 2019 Jun;143(6):e20181578. [doi: 10.1542/peds.2018-1578] [Medline: 31138669]
4. Johnston LD, O'Malley PM, et al. Monitoring the future" national survey results on drug use, 1975–2017: 2017 overview, key findings on adolescent drug use. : Institute for Social Research, University of Michigan; 2017 URL: https://monitoringthefuture.org/wp-content/uploads/2022/08/mtf-overview2017.pdf [accessed 2026-02-16]
5. Groenewald CB, Zhou C, Palermo TM, Van Cleve WC. Associations between opioid prescribing patterns and overdose among privately insured adolescents. Pediatrics 2019 Nov;144(5):e20184070. [doi: 10.1542/peds.2018-4070] [Medline: 31575622]
6. Battaglia M, Groenewald CB, Campbell F, et al. We need to talk: the urgent conversation on chronic pain, mental health, prescribing patterns and the opioid crisis. J Psychopharmacol 2023 May;37(5):437-448. [doi: 10.1177/02698811221144635] [Medline: 37171242]
7. Breglio AM, Fitzgerald TN, Moore CB, Einhorn LM. Evaluation of analgesic practice changes following the Nuss procedure in pediatric patients. J Surg Res 2023 Nov;291:289-295. [doi: 10.1016/j.jss.2023.06.018] [Medline: 37481964]
8. Densmore JC, Peterson DB, Stahovic LL, et al. Initial surgical and pain management outcomes after Nuss procedure. J Pediatr Surg 2010 Sep;45(9):1767-1771. [doi: 10.1016/j.jpedsurg.2010.01.028] [Medline: 20850618]
9. Lai K, Notrica DM, McMahon LE, et al. Cryoablation in 350 Nuss procedures: evolution of hospital length of stay and opioid use. J Pediatr Surg 2023 Aug;58(8):1435-1439. [doi: 10.1016/j.jpedsurg.2022.10.051] [Medline: 36494205]
10. Walter CM, Lee CS, Moore DL, et al. Retrospective study comparing outcomes of multimodal epidural and erector spinae catheter pain protocols after pectus surgery. J Pediatr Surg 2023 Mar;58(3):397-404. [doi: 10.1016/j.jpedsurg.2022.06.017] [Medline: 35907711]
11. Tashjian VC, Mosadeghi S, Howard AR, et al. Virtual reality for management of pain in hospitalized patients: results of a controlled trial. JMIR Ment Health 2017 Mar 29;4(1):e9. [doi: 10.2196/mental.7387] [Medline: 28356241]
12. Simonetti V, Tomietto M, Comparcini D, Vankova N, Marcelli S, Cicolini G. Effectiveness of virtual reality in the management of paediatric anxiety during the peri-operative period: a systematic review and meta-analysis. Int J Nurs Stud 2022 Jan;125:104115. [doi: 10.1016/j.ijnurstu.2021.104115] [Medline: 34781118]
13. Olbrecht VA, O'Conor KT, Williams SE, et al. Transient reductions in postoperative pain and anxiety with the use of virtual reality in children. Pain Med 2021 Nov 26;22(11):2426-2435. [doi: 10.1093/pm/pnab209] [Medline: 34175959]

14. Melzack R, Wall PD. Pain mechanisms: a new theory. Science 1965 Nov 19;150(3699):971-979. [doi: 10.1126/science.150.3699.971] [Medline: 5320816]
15. Smith KL, Wang Y, Colloca L. Impact of virtual reality technology on pain and anxiety in pediatric burn patients: a systematic review and meta-analysis. Front Virtual Real 2022 Jan;2:2022. [doi: 10.3389/frvir.2021.751735] [Medline: 36247202]
16. Indovina P, Barone D, Gallo L, Chirico A, De Pietro G, Giordano A. Virtual reality as a distraction intervention to relieve pain and distress during medical procedures: a comprehensive literature review. Clin J Pain 2018 Sep;34(9):858-877. [doi: 10.1097/AJP.0000000000000599] [Medline: 29485536]
17. Ahmadpour N, Randall H, Choksi H, Gao A, Vaughan C, Poronnik P. Virtual reality interventions for acute and chronic pain management. Int J Biochem Cell Biol 2019 Sep;114:105568. [doi: 10.1016/j.biocel.2019.105568] [Medline: 31306747]
18. Hadjiat Y, Marchand S. Virtual reality and the mediation of acute and chronic pain in adult and pediatric populations: research developments. Front Pain Res (Lausanne) 2022;3:840921. [doi: 10.3389/fpain.2022.840921] [Medline: 35599969]
19. Sánchez-Caballero E, Ortega-Donaire L, Sanz-Martos S. Immersive virtual reality for pain and anxiety management associated with medical procedures in children and adolescents: a systematic review. Children (Basel) 2024 Aug 13;11(8):975. [doi: 10.3390/children11080975] [Medline: 39201910]
20. Hoffman HG, Rodriguez RA, Gonzalez M, et al. Immersive virtual reality as an adjunctive non-opioid analgesic for pre-dominantly Latin American children with large severe burn wounds during burn wound cleaning in the intensive care unit: a pilot study. Front Hum Neurosci 2019;13(262):262. [doi: 10.3389/fnhum.2019.00262] [Medline: 31440148]
21. Olbrecht VA, O'Conor KT, Williams SE, et al. Guided relaxation–based virtual reality for acute postoperative pain and anxiety in a pediatric population: pilot observational study. J Med Internet Res 2021 Jul 12;23(7):e26328. [doi: 10.2196/26328] [Medline: 34048358]
22. Hemington KS, Cheng JC, Bosma RL, Rogachov A, Kim JA, Davis KD. Beyond negative pain-related psychological factors: resilience is related to lower pain affect in healthy adults. J Pain 2017 Sep;18(9):1117-1128. [doi: 10.1016/j.jpain.2017.04.009] [Medline: 28506777]
23. Agoston AM, Sieberg CB. Nonpharmacologic treatment of pain. Semin Pediatr Neurol 2016 Aug;23(3):220-223. [doi: 10.1016/j.spen.2016.10.005] [Medline: 27989329]
24. Turgut A, Özcan İlçe A, Öztürk H. The effect of immersive virtual reality application on anxiety, pain, and parental satisfaction in the perioperative process of children: a randomized controlled trial. Pain Manag Nurs 2024 Dec;25(6):584-590. [doi: 10.1016/j.pmn.2024.06.002] [Medline: 38955550]
25. Orgil Z, Johnson L, Karthic A, et al. Feasibility and acceptability of perioperative application of biofeedback-based virtual reality versus active control for pain and anxiety in children and adolescents undergoing surgery: protocol for a pilot randomised controlled trial. BMJ Open 2023 Jan 25;13(1):e071274. [doi: 10.1136/bmjopen-2022-071274] [Medline: 36697053]
26. Olbrecht VA, Williams SE, O'Conor KT, et al. Guided relaxation-based virtual reality versus distraction-based virtual reality or passive control for postoperative pain management in children and adolescents undergoing Nuss repair of pectus excavatum: protocol for a prospective, randomised, controlled trial (FOREVR Peds trial). BMJ Open 2020 Dec 30;10(12):e040295. [doi: 10.1136/bmjopen-2020-040295] [Medline: 33380482]
27. Chan AW, Tetzlaff JM, Gøtzsche PC, et al. SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials. BMJ 2013 Jan 8;346:e7586. [doi: 10.1136/bmj.e7586] [Medline: 23303884]
28. Hopewell S, Chan AW, Collins GS, et al. CONSORT 2025 statement: updated guideline for reporting randomized trials. JAMA 2025 Jun 10;333(22):1998-2005. [doi: 10.1001/jama.2025.4347] [Medline: 40228499]
29. Tsze DS, von Baeyer CL, Pahalyants V, Dayan PS. Validity and reliability of the verbal numerical rating scale for children aged 4 to 17 years with acute pain. Ann Emerg Med 2018 Jun;71(6):691-702. [doi: 10.1016/j.annemergmed.2017.09.009] [Medline: 29107409]
30. Price TJ, Ray PR. Recent advances toward understanding the mysteries of the acute to chronic pain transition. Curr Opin Physiol 2019 Oct;11:42-50. [doi: 10.1016/j.cophys.2019.05.015] [Medline: 32322780]
31. Pielech M, Ryan M, Logan D, Kaczynski K, White MT, Simons LE. Pain catastrophizing in children with chronic pain and their parents: proposed clinical reference points and reexamination of the Pain Catastrophizing Scale measure. Pain 2014 Nov;155(11):2360-2367. [doi: 10.1016/j.pain.2014.08.035] [Medline: 25180013]
32. Silverman WK, Fleisig W, Rabian B, Peterson RA. Childhood anxiety sensitivity index. J Clin Child Psychol 1991 Jun;20(2):162-168. [doi: 10.1207/s15374424jccp2002_7]
33. Gold JI, Mahrer NE. Is virtual reality ready for prime time in the medical space? A randomized control trial of pediatric virtual reality for acute procedural pain management. J Pediatr Psychol 2018 Apr 1;43(3):266-275. [doi: 10.1093/jpepsy/jsx129] [Medline: 29053848]
34. Proczkowska M, Ericsson E. Validity of the modified-distraction-short-scale and verbal-numeric-anxiety-fear-rating-scale for children in a preoperative setting. Paediatr Anaesth 2024 Feb;34(2):121-129. [doi: 10.1111/pan.14765] [Medline: 37728169]

35. Adams MCB, Sward KA, Perkins ML, Hurley RW. Standardizing research methods for opioid dose comparison: the NIH HEAL morphine milligram equivalent calculator. Pain 2025 Feb 3;166(8):1729-1737. [doi: 10.1097/j.pain.0000000000003529] [Medline: 39907478]
36. Morphine milligram equivalent (MME) calculator: frequently asked questions (FAQs) for HEAL researchers. : National Institutes of Health; 2025 URL: https://www.nih.gov/sites/default/files/2026-01/MME%20Calculator%20FAQ_508.pdf [accessed 2026-02-16]
37. Pitt JB, Zeineddin S, Carter M, et al. Demographics of anterior chest wall deformity patients: a tertiary children's hospital experience. J Surg Res 2024 Jan;293:451-457. [doi: 10.1016/j.jss.2023.09.040] [Medline: 37827024]
38. Mack SJ, Till BM, Huang C, et al. National trends in pectus excavatum repair: patient age, facility volume, and outcomes. J Thorac Dis 2022 Apr;14(4):952-961. [doi: 10.21037/jtd-21-1671] [Medline: 35572899]
39. Rowbotham MC. What is a "clinically meaningful" reduction in pain? Pain 2001 Nov;94(2):131-132. [doi: 10.1016/S0304-3959(01)00371-2] [Medline: 11690725]
40. Sullivan MJ. The pain catastrophizing scale: user manual. : McGill University; 2009 URL: https://drsandler.in/PCSmanual.pdf [accessed 2026-02-16]
41. Nugent SM, Lovejoy TI, Shull S, Dobscha SK, Morasco BJ. Associations of pain numeric rating scale scores collected during usual care with research administered patient reported pain outcomes. Pain Med 2021 Oct 8;22(10):2235-2241. [doi: 10.1093/pm/pnab110] [Medline: 33749760]
42. Gan TJ. Poorly controlled postoperative pain: prevalence, consequences, and prevention. J Pain Res 2017;10:2287-2298. [doi: 10.2147/JPR.S144066] [Medline: 29026331]
43. Rabbitts JA, Fisher E, Rosenbloom BN, Palermo TM. Prevalence and predictors of chronic postsurgical pain in children: a systematic review and meta-analysis. J Pain 2017 Jun;18(6):605-614. [doi: 10.1016/j.jpain.2017.03.007] [Medline: 28363861]
44. Harbaugh CM, Lee JS, Hu HM, et al. Persistent opioid use among pediatric patients after surgery. Pediatrics 2018 Jan;141(1):e20172439. [doi: 10.1542/peds.2017-2439] [Medline: 29203521]
45. Chen SF, Wang HH, Yang HY, Chung UL. Effect of relaxation with guided imagery on the physical and psychological symptoms of breast cancer patients undergoing chemotherapy. Iran Red Crescent Med J 2015 Nov;17(11):e31277. [doi: 10.5812/ircmj.31277] [Medline: 26734485]
46. Dumoulin S, Bouchard S, Ellis J, et al. A randomized controlled trial on the use of virtual reality for needle-related procedures in children and adolescents in the emergency department. Games Health J 2019 Aug;8(4):285-293. [doi: 10.1089/g4h.2018.0111] [Medline: 31135178]
47. Iannicelli AM, Vito D, Dodaro CA, et al. Does virtual reality reduce pain in pediatric patients? A systematic review. Ital J Pediatr 2019 Dec 30;45(1):171. [doi: 10.1186/s13052-019-0757-0] [Medline: 31888710]
48. Eijlers R, Utens EMWJ, Staals LM, et al. Systematic review and meta-analysis of virtual reality in pediatrics: effects on pain and anxiety. Anesth Analg 2019 Nov;129(5):1344-1353. [doi: 10.1213/ANE.0000000000004165] [Medline: 31136330]
49. Czech O, Wrzeciono A, Rutkowska A, Guzik A, Kiper P, Rutkowski S. Virtual reality interventions for needle-related procedural pain, fear and anxiety-a systematic review and meta-analysis. J Clin Med 2021 Jul 23;10(15):3248. [doi: 10.3390/jcm10153248] [Medline: 34362032]
50. Walther-Larsen S, Petersen T, Friis SM, Aagaard G, Drivenes B, Opstrup P. Immersive virtual reality for pediatric procedural pain: a randomized clinical trial. Hosp Pediatr 2019 Jul;9(7):501-507. [doi: 10.1542/hpeds.2018-0249] [Medline: 31160472]
51. Jenabi E, Bashirian S, Salehi AM, Rafiee M, Bashirian M. Virtual reality for pain reduction during intravenous injection in pediatrics: a systematic review and meta-analysis of controlled clinical trials. Clin Exp Pediatr 2023 Dec;66(12):533-537. [doi: 10.3345/cep.2022.01193] [Medline: 37321586]
52. Addab S, Hamdy R, Thorstad K, Le May S, Tsimicalis A. Use of virtual reality in managing paediatric procedural pain and anxiety: an integrative literature review. J Clin Nurs 2022 Nov;31(21-22):3032-3059. [doi: 10.1111/jocn.16217] [Medline: 35068011]
53. Alshatrat SM, Sabarini JM, Hammouri HM, Al-Bakri IA, Al-Omari WM. Effect of immersive virtual reality on pain in different dental procedures in children: a pilot study. Int J Paediatr Dent 2022 Mar;32(2):264-272. [doi: 10.1111/ipd.12851] [Medline: 34133809]
54. Gao Y, Xu Y, Liu N, Fan L. Effectiveness of virtual reality intervention on reducing the pain, anxiety and fear of needle-related procedures in paediatric patients: a systematic review and meta-analysis. J Adv Nurs 2023 Jan;79(1):15-30. [doi: 10.1111/jan.15473] [Medline: 36330583]
55. Camacho-Cruz J, Palacios-Ariza MA, Orrego-Celestino L, et al. Effectiveness of non-immersive virtual reality in the management of procedure-related pain in preschool children: a randomized clinical trial. Eur J Pediatr 2023 Sep;182(9):4103-4112. [doi: 10.1007/s00431-023-05070-5] [Medline: 37407711]
56. Mosso JL, Rizzo S, Wiederhold B, et al. Cybertherapy--new applications for discomfort reductions. Surgical care unit of heart, neonatology care unit, transplant kidney care unit, delivery room-cesarean surgery and ambulatory surgery, 27 case reports. Stud Health Technol Inform 2007;125:334-336. [Medline: 17377297]
57. Ding L, Hua H, Zhu H, et al. Effects of virtual reality on relieving postoperative pain in surgical patients: a systematic review and meta-analysis. Int J Surg 2020 Oct;82:87-94. [doi: 10.1016/j.ijsu.2020.08.033] [Medline: 32882400]

58. Eijlers R, Dierckx B, Staals LM, et al. Virtual reality exposure before elective day care surgery to reduce anxiety and pain in children: a randomised controlled trial. Eur J Anaesthesiol 2019 Oct;36(10):728-737. [doi: 10.1097/EJA.0000000000001059] [Medline: 31356373]
59. Vagnoli L, Bettini A, Amore E, De Masi S, Messeri A. Relaxation-guided imagery reduces perioperative anxiety and pain in children: a randomized study. Eur J Pediatr 2019 Jun;178(6):913-921. [doi: 10.1007/s00431-019-03376-x] [Medline: 30944985]
60. Bernaerts S, Bonroy B, Daems J, et al. Virtual reality for distraction and relaxation in a pediatric hospital setting: an interventional study with a mixed-methods design. Front Digit Health 2022;4:866119. [doi: 10.3389/fdgth.2022.866119] [Medline: 35712230]
61. Hoag JA, Karst JS, Bingen KM, Palou-Torres A, Yan K. Distracting through procedural pain and distress using virtual reality and guided imagery in pediatric, adolescent, and young adult patients: randomized controlled trial. J Med Internet Res 2022 Apr 18;24(4):e30260. [doi: 10.2196/30260] [Medline: 35436209]
62. Xiang H, Shen J, Wheeler KK, et al. Efficacy of smartphone active and passive virtual reality distraction vs standard care on burn pain among pediatric patients: a randomized clinical trial. JAMA Netw Open 2021 Jun 1;4(6):e2112082. [doi: 10.1001/jamanetworkopen.2021.12082] [Medline: 34152420]
63. Gold JI, SooHoo M, Laikin AM, Lane AS, Klein MJ. Effect of an immersive virtual reality intervention on pain and anxiety associated with peripheral intravenous catheter placement in the pediatric setting: a randomized clinical trial. JAMA Netw Open 2021 Aug 2;4(8):e2122569. [doi: 10.1001/jamanetworkopen.2021.22569] [Medline: 34432011]

## Abbreviations

**360-V:** 360° video
**AUC:** area under the curve
**CCHMC :** Cincinnati Children's Hospital Medical Center
**CONSORT:** Consolidated Standard of Reporting Trials
**MME:** morphine milligram equivalents
**NRS:** numerical rating scale
**PCS :** Pain Catastrophizing Scale
**POD :** postoperative day
**REDCap:** Research Electronic Data Capture
**SPIRIT:** Standard Protocol Items: Recommendations for Interventional Trials
**VR:** virtual reality
**VR-D :** distraction-based gaming virtual reality
**VR-DGR:** distraction-based guided relaxation virtual reality
**VR-GR:** guided relaxation–based virtual reality

*Edited by N Rohatgi; submitted 18.Jul.2025; peer-reviewed by I Cornelisz, JW Gander; revised version received 20.Jan.2026; accepted 06.Feb.2026; published 10.Mar.2026.*

*Please cite as:*
*Walter CM, Froass D, Bell N, Haack L, Boehmer C, Bruguera Torres C, Spivak R, Chou M, Geisler K, O'Conor K, Williams SE, Ding L, King CD, Olbrecht VA*
*Virtual Reality for the Management of Postoperative Pain and Anxiety in Children and Adolescents Undergoing Nuss Repair of Pectus Excavatum: Randomized Controlled Trial*
*JMIR Perioper Med 2026;9:e80902*
*URL: https://periop.jmir.org/2026/1/e80902*
*doi:10.2196/80902*

# Evaluating the Impact of Virtual Reality on Orthopedic Trauma Skills Acquisition Among Surgical Residents: Randomized Crossover Study

Sandipan Chatterjee[1], PhD; Khairul Faizi Mohammad[2], MD; Monica Ghidinelli[1], PhD

[1]AO Education Institute, AO Foundation, Clavadelerstrasse 8, Davos, Switzerland

[2]Department of Orthopedics, Pantai Hospital Cheras, Kuala Lumpur, Malaysia

**Corresponding Author:**
Sandipan Chatterjee, PhD
AO Education Institute, AO Foundation, Clavadelerstrasse 8, Davos, Switzerland

## Abstract

**Background:** Orthopedic trauma skills training is time-consuming and expensive. Current training modalities rely heavily on synthetic bone models, anatomical laboratory simulations, or assistance in surgeries (the apprenticeship model). Virtual reality (VR) appears to present a promising complement to current training modalities.

**Objective:** This study evaluated the effectiveness of VR training on surgical performance and gauged learning preferences among orthopedic trauma residents in Malaysia.

**Methods:** In total, 123 orthopedic residents were randomly assigned to 2 groups. One group practiced for about 30 minutes using VR glasses, followed by conventional nailing exercises on synthetic bones, while the other group first performed the nailing exercise, followed by VR practice. Performance was measured by time to completion of the exercise, and participants completed a postexercise survey.

**Results:** Participants who completed VR training before the synthetic bone nailing exercise were significantly faster, completing the task between 4 ($P$=.05) and 7 ($P$=.002) minutes more quickly than without VR training. In addition, VR training improved self-assessed performance during the exercise. Survey data revealed that while 43% (50/117) of participants preferred conventional methods of learning (lectures, discussions, and hands-on simulations), 89% (104/117) of participants supported VR use as an adjunct to conventional methods of learning. Less than 2% (2/117, 1.7%) of participants indicated that conventional methods of learning were outdated.

**Conclusions:** A single session of VR training significantly reduced completion times and improved self-assessment of competence in orthopedic trauma simulation exercises. Although learners continue to value conventional training modalities, there is a strong desire to include VR as a supplementary tool. Its integration into surgical curricula may accelerate skill acquisition, especially in low-resource settings with limited access to high-fidelity simulation labs. In addition, the availability of VR training modules in hospitals could help residents and junior consultants prepare for surgery.

***(JMIR Perioper Med 2026;9:e79343)*** doi:10.2196/79343

**KEYWORDS**

virtual reality; orthopedic procedures; graduate medical education; simulation training; fracture fixation

## Introduction

### Background

The growing complexity of orthopedic trauma surgery has heightened the demand for advanced and effective training modalities. Conventional approaches such as synthetic bone models, anatomical laboratory simulations, and the apprenticeship model of learning through assisting in surgeries are time-consuming, expensive, and limited in availability [1,2]. Anatomical laboratory simulations provide high anatomical fidelity and tissue realism but are logistically demanding (eg, cadaver procurement, facility access, higher per-learner cost, and lower frequency). These constraints, particularly pronounced in low-resource settings, limit sufficient exposure to trauma case variations and present barriers to skill acquisition.

### Prior Work

Alternatives are being explored to complement traditional training modalities. Virtual reality (VR) has emerged as a promising tool that offers immersive, repeatable, and cost-effective simulations in a controlled environment [3-5]. In contrast to anatomical laboratories, immersive VR has standardized scenarios, providing immediate feedback and repetition at scale. VR also offers a lower marginal cost and greater scheduling flexibility, albeit with reduced tactile realism

and haptic resistance compared with cadaveric tissue. Recent studies have highlighted the benefits of VR for orthopedic surgery, where procedural repetition and real-time feedback, which are critical for skill acquisition, can help reduce the learning curve [6]. VR in orthopedic training appears particularly effective in simulating procedures such as intramedullary nailing, arthroscopy, and arthroplasty [7-10]. Beyond technical skill acquisition, VR can also support the development of nontechnical skills, including decision-making and teamwork—critical elements in high-stakes surgical environments [11,12].

One of VR's primary advantages is its scalability, particularly in low-resource settings [13]. The increasing availability of VR platforms such as Oculus Rift and HTC Vive has the potential to increase training opportunities in such regions [5]. The ability to simulate a broad range of clinical scenarios, including rare or complex cases, provides a comprehensive training experience that would otherwise be difficult to achieve through traditional methods.

Although prior studies have established VR's general utility in surgical training, our study is the first to observe immediate skill transfer after a single VR session for both tibial and femoral nailing procedures, in a Malaysian residency context. Additionally, we examined the effect of VR training on self-assessed performance and gauged learner preferences for different training modalities.

### Training Challenges in Malaysia

In Malaysia, orthopedic residency is a 4-year training program under the Ministry of Health and the Ministry of Education. Upon graduation, residents serve in hospitals run by either the Ministry of Health or the Ministry of Education. Selection is based on open entrance examinations and interviews [14].

Surgical training in Malaysia is generally considered "in work training." Hence, the level of training gained depends on the level provided by the trainer at the hospital in which the candidate is serving. The training level also varies depending on the opportunity to perform surgery and the types of cases available within a particular hospital. Other constraints particular to the region are the lack of sufficient hospital- or university-based courses. Anatomical laboratories are available but are expensive and infrequent [15].

According to a report published by the University of Malaya, there were an estimated 900 orthopedic surgeons in Malaysia in 2018, giving a ratio of 1 orthopedic surgeon per 40,000 population [14]. To meet the suggested World Health Organization ratio of 1:30,000 population, orthopedics in Malaysia faces the challenge of increasing surgeon numbers accordingly. Although approximately 70 residents are recruited annually and placed at 30 accredited training centers, they must also navigate economic realities to provide the best treatment within the limitations of public funding and patient affordability [14].

### Goal of This Study

This study aimed to evaluate the impact of VR-based training on surgical performance and to explore learning preferences of orthopedic trauma residents in Malaysia.

## Methods

### Educational Intervention

The educational intervention was integrated into a 3-day course that combined lectures, small group discussions, and hands-on practical exercises. Participants practiced a common tibia nailing procedure and a femoral nailing procedure on both VR headsets and synthetic bones. VR training was conducted using Oculus Rift S headsets (Meta Platforms Inc), running Johnson & Johnson MedTech simulation software. The headset recorded procedure completion time, procedural steps, and errors, where applicable.

Each practical exercise workstation had 3 participants per synthetic bone. For VR training, participants practiced the nailing procedure once. In total, 30 VR headsets were available for this study, divided across 6 rooms, each with 5 participants, 1 faculty member, 1 laptop, and 1 technical staff member. The rooms were divided into individual grids of 1.5 $m^2$ (Figure 1). The hotel Wi-Fi network was used to connect the headsets.

**Figure 1.** Setup of the virtual reality training session. Consent for using this image was obtained via course registration.

### Tibia Nailing Exercise

The tibia nail used for the synthetic bone exercise was the Expert Tibia Nail (Johnson & Johnson). For the VR module, the Tibia Nail Advanced (Johnson & Johnson) module was used. As the principal and major steps of both nailing systems are similar, they were deemed equivalent for comparison in this study. The tibia nailing procedure was performed during an AO (Arbeitsgemeinschaft für Osteosynthesefragen) Trauma Basic Principles course (cohort 1).

### Femoral Nailing Exercise

The femoral nail used for the synthetic bone exercise was the Trochanteric Fixation Nail Advanced Proximal Femoral Nailing System (Johnson & Johnson). The same nailing procedure was used for the VR module. The femoral nailing procedure was performed during the AO Trauma Basic Principles course (cohort 1) and the AO Trauma Advanced Principles course (cohort 2).

### Study Design

We used a 2-period crossover design. Participants were randomized to one of two sequences: (1) VR practice followed by the synthetic bone exercise or (2) the synthetic bone exercise followed by VR practice. After completing the first period, participants crossed over to the other modality. The primary end point was time (in minutes) to complete the synthetic bone exercise, with or without prior VR practice. These comparisons are reported separately for tibial and femoral procedures. Completion times for both modalities (VR and synthetic bone) are reported.

A qualitative postexercise survey assessed self-perceived performance and learning preferences. The questionnaire was developed, revised, and pilot-tested with 5 surgeons and administered via Microsoft Forms (version 16.0; Microsoft Corp).

### Population

Participants in this study were those who attended the AO Trauma Basic and Advanced Principles courses on the Principles of Fracture Management in Malaysia in March 2023. Beyond course attendance and consent, no inclusion or exclusion criteria were applied. All participants were able to complete VR training, as no cases of nausea or dizziness were observed while using the VR headsets.

### Data Collection

Participant demographic data were collected at baseline. For the synthetic bone exercises, each participant captured and reported the synthetic bone exercise completion time using a stopwatch. The completion times of the VR module were automatically captured via the VR headset software. Data were collected after each exercise and analyzed using Microsoft Excel (version 16.0; Microsoft Corp). Survey responses were collected after the exercise through Microsoft Forms.

### Analysis

A descriptive analysis was performed to determine the frequencies and percentages of the response selections. The tables and figures display frequency counts and/or percentages for each response option. A paired 2-tailed *t* test was used to test the effect of VR training on time to completion of the synthetic bone practical exercise. A Bonferroni post hoc test was conducted to correct for multiple testing.

### Ethical Considerations

According to the ethics committee of the Canton of Zurich, this study did not require ethical committee authorization (Req-2025-00605). Informed consent was obtained from participants prior to the start of the course and no compensation was provided to study participants for taking part in this study. In the survey, we included the following statement of purpose, which disclosed our intended use of the data: "The information you provide will be anonymized and made available to the study group in aggregate form. The data will be used for research purposes."

## Results

### Participant Demographics

This study included 2 cohorts with 63 and 60 participants from the AO Trauma Basic and Advanced Principles courses, respectively (Table 1). The average age between these two cohorts was significantly different (*P*<.001), as was the average years of orthopedic trauma training (*P*<.001) (Table 1). There was no significant difference in the number of femoral nailing cases performed by the participants of each cohort.

**Table .** Participant demographics and experience level.[a]

| Participant demographics | Cohort 1[b] | Cohort 2[c] | *P* value |
|---|---|---|---|
| Sex, n (%) | | | —[d] |
| Female | 15 (24) | 5 (8) | |
| Male | 48 (76) | 55 (92) | |
| Age (years), mean (SD) | 33.0 (3.6) | 37.36 (2.8) | *<.001* |
| Career stage, n (%) | | | —[d] |
| Specialist | 14 (22) | 57 (95) | |
| Master's trainee | 9 (14) | 2 (3) | |
| Service medical officer (resident) | 40 (64) | 1 (2) | |
| Experience, mean (SD) | | | *<.001* |
| Duration of orthopedic training (years) | 4.5 (3.2) | 9.6 (3.3) | |
| Number of tibia nailing cases performed as a primary surgeon | 10 (12) | — | |
| Number of tibia nailing cases performed as an assistant surgeon | 15 (19) | — | |
| Number of femoral nailing cases performed as a primary surgeon | 10 (13) | 15 (20) | |
| Number of femoral nailing cases performed as an assistant surgeon | 13 (16) | 16 (24) | |

[a]Averages include SDs, and *P* values when significant are italicized.

[b]Cohort 1: n=63.

[c]Cohort 2: n=60.

[d]Not applicable.

### Completion Time for Tibia Nailing Exercise

For each procedure, we compared the synthetic bone completion time among participants who performed VR practice before the synthetic bone task versus those who performed it after. For cohort 1, the survey response rates (proportion of participants who completed the postexercise survey) varied from 80% (50/63) to 92% (58/63). For cohort 2, the response rate was approximately 93% (55/60).

For the tibia nailing exercise, participants who trained on the VR module prior to performing the synthetic bone practical exercise were, on average, 5 minutes faster in completing the nailing exercise (P=.007; Table 2). A significant decrease in time needed to finish the VR module was also noted when participants first completed the synthetic bone practical exercise (*P*=.02; Table 2).

**Table .** Completion times for the tibia nailing exercise.

| Cohort 1 | VR[a] module then synthetic bone exercise, mean (95% CI) | Synthetic bone exercise then VR module, mean (95% CI) | *P* [b] *value* |
|---|---|---|---|
| Time (min) to complete synthetic bone exercise | 18.48 (16.19-20.77) | 23.43 (20.8-26.01) | *.007* |
| Time (min) to complete VR module | 34.97 (31.51-38.44) | 29.82 (27.43-32.2) | *.02* |

[a]VR: virtual reality.

[b]Italicized *P* values are significant.

## Completion Time for Femoral Nailing Exercise

Participants from cohort 1 who trained on the VR module prior to performing the synthetic bone practical exercise were, on average, approximately 7 minutes faster in completing the femoral nailing exercise (*P*<.01; Table 3). No significant difference in the VR module completion time was noted after participants performed the synthetic bone exercise (*P*=.09; Table 3). Similarly, in cohort 2, participants were on average approximately 4 minutes faster in completing the synthetic bone practical exercise after training on the VR module (*P*=.05; Table 3), and no significant difference in the VR module completion time was noted (*P*=.54; Table 3).

**Table .** Completion times for the femoral nailing exercise.

| Cohort and outcome | VR[a] module then synthetic bone exercise, mean (95% CI) | Synthetic bone exercise then VR module, mean (95% CI) | *P*[b] value |
|---|---|---|---|
| Cohort 1 | | | |
| Time (min) to complete synthetic bone exercise | 20.71 (18.11-23.32) | 27.55 (24.25-30.86) | *<.01* |
| Time (min) to complete VR module | 18.65 (16.82-20.48) | 16.69 (15.36-18.01) | .09 |
| Cohort 2 | | | |
| Time (min) to complete synthetic bone exercise | 16.51 (15.03-17.98) | 20.73 (16.89-24.58) | *.05* |
| Time (min) to complete VR module | 21.57 (19.65-23.49) | 22.45 (20.42-24.48) | .54 |

[a]VR: virtual reality.

[b]Italicized *P* values are significant.

## Impact of VR Training on Self-Assessed Performance

Analysis of the qualitative questionnaire showed that after VR training, a higher number of participants rated the ease of understanding the nailing practical exercise in the “extremely easy” and/or “very easy” categories. This trend (results not statistically significant) was also noticed when participants were asked to evaluate their performance during the synthetic bone exercise (with the exception of cohort 1 for femoral nailing). This observation was seen for both the tibial and femoral nailing exercises (Table 4 and Table 5).

**Table .** Impact of virtual reality (VR) training on understanding and performance during the tibia synthetic bone nailing exercise.

| Tibia nailing outcomes | Synthetic bone exercise without VR training, n (%) | Synthetic bone exercise with VR training, n (%) |
|---|---|---|
| How would you rate the ease of understanding the practical? | | |
| Extremely easy | 0 (0)[a] | 4 (13)[b] |
| Very easy | 9 (30)[c] | 10 (33)[d] |
| Easy | 17 (57)[d] | 10 (33)[d] |
| Somewhat easy | 4 (13)[b] | 6 (20)[c] |
| Not at all easy | 0 (0)[a] | 0[a] (0) |
| How would you rate your performance in this practical? | | |
| Excellent | 2 (7)[b] | 4 (13)[b] |
| Very good | 9 (30)[c] | 13 (43)[d] |
| Good | 19 (63)[d] | 13 (43)[d] |
| Poor | 0[a] (0) | 0[a] (0) |
| Very poor | 0[a] (0) | 0[a] (0) |

[a]Lowest frequency of responses.

[b]Moderate frequency of responses.

[c]High frequency of responses.

[d]Highest frequency of responses.

**Table .** Impact of prior virtual reality (VR) training on understanding and performance during the femoral synthetic bone nailing exercise.

| Femoral nailing outcomes | Cohort 1 | | Cohort 2 | |
|---|---|---|---|---|
| | Synthetic bone without VR training, n (%) | Synthetic bone with VR training, n (%) | Synthetic bone without VR training, n (%) | Synthetic bone with VR training, n (%) |
| How would you rate the ease of understanding the practical? | | | | |
| Extremely easy | 0 (0)[a] | 4 (13)[b] | 1 (3)[a] | 4 (15)[b] |
| Very easy | 3 (10)[c] | 5 (16)[b] | 8 (29)[c] | 9 (33)[c] |
| Easy | 21 (73)[d] | 13 (42)[d] | 13 (46)[d] | 13 (48)[d] |
| Somewhat easy | 3 (10)[c] | 8 (26)[c] | 5 (18)[b] | 1 (4)[a] |
| Not at all easy | 2[b] (7) | 1 (3)[a] | 1 (3)[a] | 0 (0)[a] |
| How would you rate your performance in this practical? | | | | |
| Excellent | 0[a] (0) | 2 (7)[b] | 5 (18)[c] | 4 (15)[b] |
| Very good | 10 (35)[c] | 7 (23)[c] | 3 (11)[b] | 10 (37)[c] |
| Good | 16 (55)[d] | 19 (61)[d] | 18 (64)[d] | 12 (44)[d] |
| Poor | 3 (10)[b] | 2 (7)[b] | 2 (7)[b] | 1 (4)[b] |
| Very poor | 0[a] (0) | 1 (3)[a] | 0 (0)[a] | 0 (0)[a] |

[a]Lowest frequency of responses.

[b]Moderate frequency of responses.

[c]High frequency of responses.

[d]Highest frequency of responses.

### Learning Preferences

When asked to compare VR with conventional methods of learning (lectures and synthetic bone practical exercises), participants were strongly in favor of VR as a supplement to conventional methods of learning. Only 3.4% (4/117) chose VR modules alone as their preferred method of learning, 89% (104/117) preferred VR as an adjunct to conventional methods, and only 8% (9/117) believed VR added no value (Table 6).

When asked about the reverse, evaluating conventional methods of learning versus VR, 43% (50/117) preferred conventional methods of learning, 56% (65/117) preferred conventional methods as an adjunct to VR, and less than 2% (2/117, 1.7%) considered conventional methods to be outdated (Table 6).

**Table .** Participant learning preferences.

| Learning preference | Participants, n (%) |
|---|---|
| How do you rate the VR[a] module vs conventional methods of learning (lectures and hands-on training with synthetic bones)? | |
| Prefer the VR module | 4 (3.42) |
| Prefer VR as an adjunct to conventional methods of learning | 104 (88.89) |
| VR does not add to conventional methods of learning | 9 (7.69) |
| How do you rate conventional methods of learning (lectures and hands-on training with synthetic bones) vs the VR module? | |
| Prefer conventional methods of learning | 50 (42.74) |
| Prefer conventional methods of learning as an adjunct to VR teaching | 65 (55.56) |
| Conventional methods of learning are outdated | 2 (1.71) |

[a]VR: virtual reality.

### Association of Age and Experience Level With Completion Times of the Femoral Nailing Exercise and VR Module

As the average age of the 2 cohorts were significantly different (33.0, SD 3.6 years for cohort 1 and 37.6, SD 2.8 years for cohort 2; *P*<.001; Table 1) and the average number of years of orthopedic training was also significantly different (4.5, SD 3.2 years vs 9.6, SD 3.3 years; *P*<.001; Table 1), we conducted an unplanned subanalysis examining the influence of age and experience on completion times of the femoral nailing exercise. Participants in cohort 2 (older, with more years of training) were significantly faster in completing the synthetic bone nailing exercise than those in cohort 1 (20.7 vs 27.5 minutes; *P*=.01; Table 7). However, when comparing the VR module completion times, cohort 1 (younger, with fewer years of training) was significantly faster than cohort 2 (18.6 vs 21.8 minutes; *P*=.03; Table 7).

**Table .** Effect of age and experience on completion time of nailing exercises.

| Outcomes | Cohort 1 (time in minutes) | Cohort 2 (time in minutes) | *P*[a] value |
|---|---|---|---|
| Time to complete the femoral nailing *synthetic bone* exercise (without virtual reality training) | 27.5 | 20.73 | *.01* |
| Time to complete the femoral nailing *virtual reality module* (without prior synthetic bone exercise) | 18.65 | 21.57 | *.03* |

[a]Italicized *P* values are significant.

## Discussion

### Principal Results

This study demonstrated immediate skill transfer from a single VR session to synthetic bone performance across 2 orthopedic procedures in a Malaysian residency context.

Our primary finding suggests that VR supports rapid skill acquisition and transfer to synthetic bone models. Additionally, we explored learning preferences, which could inform future curriculum design in low-resource countries. Although traditional hands-on training remains highly valued, most participants viewed VR positively as a tool complementary to conventional training methods. Our qualitative data further support VR's effectiveness, with participants reporting a better understanding of the practical exercises and higher self-rated performance after VR training.

An interesting secondary finding emerged from our unplanned subanalysis comparing cohorts of different experience levels. The more experienced participants performed the conventional synthetic bone femoral nailing exercise significantly faster than the less-experienced group. This result confirms that conventional surgical skills improve with years of practice.

In contrast, the younger and less-experienced cohort completed the VR module significantly faster. This finding suggests that younger trainees may adapt more quickly to digital learning environments, possibly due to greater familiarity with technology. However, given baseline differences and the

pragmatic study design, the analysis of age and experience is exploratory and should not be interpreted as causal.

Several practical implications emerge from our findings. First, in the context of Malaysia's orthopedic training landscape, where there is a need to train more surgeons to meet the World Health Organization–recommended ratio of 1:30,000 population, VR offers a scalable supplement that could help standardize training while optimizing the use of limited resources. With limited access to cadaveric specimens and economic constraints, VR platforms could provide residents with additional repetitions of procedures before they enter the operating room.

Second, the demonstrated reduction in time to complete practical exercises after VR training shows the transferability of skills from VR to real-life settings. Incorporating VR into preprocedural preparation could improve operating room efficiency. This is particularly valuable in resource-constrained settings where operating room time is limited and maximizing throughput is essential.

Third, the improved self-assessment of performance after VR training indicates enhanced procedural confidence, which could translate to reduced stress and potentially better outcomes in actual surgical situations. The combination of faster performance and increased understanding suggests that VR training may help flatten the learning curve for complex surgical procedures and offers a solution that complements rather than replaces traditional training methods [16,17].

### Comparison With Prior Work

The growing body of evidence on the use of VR simulations for orthopedic surgery skills training and education has resulted in VR being included as part of the third technological wave in orthopedics [18]. Although there are still some questions concerning cost-effectiveness and skill retention using VR simulations, there is no doubt about VR's effectiveness in surgical training, particularly in orthopedics, where procedural repetition and spatial understanding are critical [19,20]. Our work aligns well with this literature, as we demonstrate immediate skill transfer after a single VR training intervention for not 1 but 2 orthopedic procedures.

A strong preference for a mixed-learning approach (conventional and VR) among participants reflects the need to rethink the role of VR in surgical skills training, especially in the context of low-resource settings [21,22], and to offer a scalable solution that complements rather than replaces traditional training methods. Acquiring VR headsets would be a 1-time investment for the training facility or could be sourced from companies willing to sponsor the event. On the other hand, skills training on synthetic bones requires the purchase of bone models for every participant and training event.

### Strengths

This study benefits from several methodological strengths. The crossover design allowed each participant to experience both training modalities, reducing individual variation as a confounding factor. The relatively large sample size (N=123) and the inclusion of 2 distinct cohorts with different experience levels enhance the generalizability of our findings. Additionally, testing 2 different nailing procedures (tibia and femoral) demonstrates that the observed effects are not procedure specific.

### Limitations

Several limitations warrant consideration when interpreting our results. First, completion times for the synthetic bone exercises were self-reported, whereas VR module completion times were captured by the software. This could introduce a reporting bias between the 2 modalities. Although we provided participants with standardized instructions, reporting errors cannot be excluded. Future studies could benefit from objective timing methods or video recording of procedures for more accurate measurement. Second, we evaluated effectiveness primarily through completion time, which, while important, represents only 1 dimension of surgical competence. Additional outcome measures such as procedural accuracy, error rates, or expert assessment of technique would provide a more comprehensive evaluation of VR's impact on skill acquisition. Third, our study assessed immediate skill transfer but did not include follow-up assessments to determine the durability of learned skills. Longitudinal studies would be valuable to assess whether VR-acquired skills persist over time and translate to improved performance in actual surgical cases. Finally, this was a pragmatic, course-embedded evaluation. Beyond course attendance and consent, we did not apply formal inclusion or exclusion criteria, which may contribute to baseline heterogeneity. Using immersive VR headsets has been shown to cause dizziness or nausea (cybersickness) [23]. Although we did not experience any such cases in this study, it needs to be considered while planning similar interventions.

## Conclusions

This study provides new evidence that even a single, brief VR training session can significantly decrease completion times of simulated nailing procedures. This observation, coupled with improved self-reported procedural understanding and performance, suggests that VR represents a valuable educational tool for orthopedic skills training.

Although conventional hands-on training remains an integral part of residency training, there is a strong preference for a mixed approach incorporating VR technologies. In the context of Malaysia's need to train more orthopedic surgeons, VR training could offer a scalable complement to traditional methods that could help standardize training while optimizing resource use. This study also highlights the need for tailored implementation strategies that account for varying levels of technological familiarity. Finally, as VR technology continues to evolve and become more accessible, its integration into orthopedic training curricula represents a promising approach to enhance surgical education.

## Acknowledgments

The authors thank the AO (Arbeitsgemeinschaft für Osteosynthesefragen) Foundation for organizing and running the courses where this study took place and the participants for agreeing to take part in this study. The authors thank Johnson & Johnson MedTech Malaysia for providing the virtual reality headsets and for in-kind support to run the course. The authors also thank the course chairperson, Dr Nazari Ahmad Tarmuzi, for making faculty available to support the study and Iris Lam for logistic support.

## Funding

Financial and logistic support for conducting the educational event was provided by the AO Foundation, through an education grant from Johnson & Johnson MedTech to the AO Foundation.

## Data Availability

All data generated and analyzed during this study are included in this published article.

## Authors' Contributions

SC and KFM designed the study and collected the data. All authors contributed to data analysis, manuscript drafting, and revision of the manuscript.

## Conflicts of Interest

None declared.

## References

1. Babineau TJ, Becker J, Gibbons G, et al. The “cost” of operative training for surgical residents. Arch Surg 2004 Apr;139(4):366-369. [doi: 10.1001/archsurg.139.4.366] [Medline: 15078701]
2. Marchand LS, Sciadini MF. Simulation training in fracture surgery. J Am Acad Orthop Surg 2020 Nov 1;28(21):e939-e947. [doi: 10.5435/JAAOS-D-20-00076] [Medline: 32796368]
3. Hasan LK, Haratian A, Kim M, Bolia IK, Weber AE, Petrigliano FA. Virtual reality in orthopedic surgery training. Adv Med Educ Pract 2021;12:1295-1301. [doi: 10.2147/AMEP.S321885] [Medline: 34785971]
4. Li T, Yan J, Gao X, et al. Using virtual reality to enhance surgical skills and engagement in orthopedic education: systematic review and meta-analysis. J Med Internet Res 2025 May 30;27:e70266. [doi: 10.2196/70266] [Medline: 40446337]
5. Ntakakis G, Plomariti C, Frantzidis C, Antoniou PE, Bamidis PD, Tsoulfas G. Exploring the use of virtual reality in surgical education. World J Transplant 2023 Feb 18;13(2):36-43. [doi: 10.5500/wjt.v13.i2.36] [Medline: 36908307]
6. Makransky G, Mayer RE. Benefits of taking a virtual field trip in immersive virtual reality: evidence for the immersion principle in multimedia learning. Educ Psychol Rev 2022;34(3):1771-1798. [doi: 10.1007/s10648-022-09675-4] [Medline: 35475019]
7. Blumstein G, Zukotynski B, Cevallos N, et al. Randomized trial of a virtual reality tool to teach surgical technique for tibial shaft fracture intramedullary nailing. J Surg Educ 2020;77(4):969-977. [doi: 10.1016/j.jsurg.2020.01.002] [Medline: 32035854]
8. Walbron P, Common H, Thomazeau H, et al. Virtual reality simulator improves the acquisition of basic arthroscopy skills in first-year orthopedic surgery residents. Orthop Traumatol Surg Res 2020 Jun;106(4):717-724. [doi: 10.1016/j.otsr.2020.03.009] [Medline: 32284277]
9. Unal YC, Adanas C, Kaya S, Ozkan S, Ismailov U. Virtual reality headset versus traditional training in orthopedic surgery: a comparative study on real surgical performance in total knee arthroplasty. Medicine (Baltimore) 2024 Nov 22;103(47):e40615. [doi: 10.1097/MD.0000000000040615] [Medline: 39809210]
10. Crockatt WK, Confino JE, Kopydlowski NJ, Jobin CM, Levine WN. Comparing skill acquisition and validity of immersive virtual reality with cadaver laboratory sessions in training for reverse total shoulder arthroplasty. JB JS Open Access 2023;8(3):e22.00141. [doi: 10.2106/JBJS.OA.22.00141] [Medline: 37415724]
11. Wan T, Liu K, Li B, Wang X. Validity of an immersive virtual reality training system for orthognathic surgical education. Front Pediatr 2023;11:1133456. [doi: 10.3389/fped.2023.1133456] [Medline: 37033170]
12. Hiemstra LA, Williams H, Sasyniuk TM, Kerslake S. Orthopaedic sport medicine surgeons and fellows value immersive virtual reality for improving surgical training, procedural planning, and distance learning. J ISAKOS 2024 Aug;9(4):769-773. [doi: 10.1016/j.jisako.2024.05.002] [Medline: 38734310]
13. Please H, Narang K, Bolton W, et al. Virtual reality technology for surgical learning: qualitative outcomes of the first virtual reality training course for emergency and essential surgery delivered by a UK-Uganda partnership. BMJ Open Qual 2024 Jan 29;13(1):e002477. [doi: 10.1136/bmjoq-2023-002477] [Medline: 38286564]
14. Orthopaedic surgery postgraduate training in Malaysia. Majlis Dekan Fakulti Perubatan Universiti Awam Malaysia. 2020. URL: https://medicine.um.edu.my/pdf/postgraduate/NPMC/NPMC_OrthopaedicSurgeryPGTraining.pdf [accessed 2026-04-18]

15. Ibrahim S. The evolution of postgraduate orthopaedic training in Malaysia. International Society of Orthopaedic Surgery and Traumatology (SICOT). 2015. URL: https://www.sicot.org/enewsletter-75-editorial [accessed 2026-04-18]
16. Seymour NE, Gallagher AG, Roman SA, et al. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. Ann Surg 2002 Oct;236(4):458-463. [doi: 10.1097/00000658-200210000-00008] [Medline: 12368674]
17. Vaishya R, Misra S, Singh LP, et al. Surgical training and learning curves in orthopedic surgery – a review. J Orthop Rep 2025 Aug;4(2):100584. [doi: 10.1016/j.jorep.2025.100584]
18. Zhang J, Ye Z. Digital orthopedics: the third technological wave of orthopedics. Intell Med 2025 May;5(2):91-94. [doi: 10.1016/j.imed.2024.09.003]
19. Lohre R, Bois AJ, Pollock JW, et al. Effectiveness of immersive virtual reality on orthopedic surgical skills and knowledge acquisition among senior surgical residents: a randomized clinical trial. JAMA Netw Open 2020 Dec 1;3(12):e2031217. [doi: 10.1001/jamanetworkopen.2020.31217] [Medline: 33369660]
20. Cevallos N, Zukotynski B, Greig D, Silva M, Thompson RM. The utility of virtual reality in orthopedic surgical training. J Surg Educ 2022;79(6):1516-1525. [doi: 10.1016/j.jsurg.2022.06.007] [Medline: 35821110]
21. Cate G, Barnes J, Cherney S, et al. Current status of virtual reality simulation education for orthopedic residents: the need for a change in focus. Global Surg Educ 2023;2(1):46. [doi: 10.1007/s44186-023-00120-w] [Medline: 38013875]
22. Heo K, Cheng S, Joos E, Joharifard S. Use of innovative technology in surgical training in resource-limited settings: a scoping review. J Surg Educ 2024 Feb;81(2):243-256. [doi: 10.1016/j.jsurg.2023.11.004] [Medline: 38161100]
23. Cossio S, Chiappinotto S, Dentice S, et al. Cybersickness and discomfort from head-mounted displays delivering fully immersive virtual reality: a systematic review. Nurse Educ Pract 2025 May;85:104376. [doi: 10.1016/j.nepr.2025.104376] [Medline: 40267853]

## Abbreviations

**AO:** Arbeitsgemeinschaft für Osteosynthesefragen
**VR:** virtual reality

*Edited by N Rohatgi; submitted 19.Jun.2025; peer-reviewed by G Rodriguez-Fuentes, OR Babalola; revised version received 11.Mar.2026; accepted 19.Mar.2026; published 06.May.2026.*

*Please cite as:*
*Chatterjee S, Mohammad KF, Ghidinelli M*
*Evaluating the Impact of Virtual Reality on Orthopedic Trauma Skills Acquisition Among Surgical Residents: Randomized Crossover Study*
*JMIR Perioper Med 2026;9:e79343*
*URL: https://periop.jmir.org/2026/1/e79343*
*doi:10.2196/79343*

# A Novel Customizable Datamart and Tableau Dashboard to Monitor Multiple Enhanced Recovery After Surgery Programs: Development and Validation Study

Sunitha Margaret Singh[1], MD; Susannah Oster[2], BA; Efrat Bolze[3], BA; Aaron Sasson[4], MD; James Nicholson[5], MD; Elliott Bennett-Guerrero[2], MD

[1]Department of Perioperative Surgical Services, Stony Brook University Medical Center, 101 Nichols Road, Stony Brook, NY, United States

[2]Department of Anesthesiology, Stony Brook University Medical Center, Stony Brook, NY, United States

[3]Enterprise Analytics, Stony Brook Medicine Information Technology (SBMIT), St. James, NY, United States

[4]Department of Surgery, Stony Brook University Medical Center, Stony Brook, NY, United States

[5]Department of Orthopedics, Stony Brook University Medical Center, Stony Brook, NY, United States

**Corresponding Author:**
Sunitha Margaret Singh, MD
Department of Perioperative Surgical Services, Stony Brook University Medical Center, 101 Nichols Road, Stony Brook, NY, United States

## Abstract

**Background:** Enhanced recovery after surgery (ERAS) programs bundle evidence-based interventions to standardize care, expedite recovery, and improve outcomes. As ERAS programs have expanded, it has become clear that a major challenge is monitoring the compliance of bundle elements and outcomes to feedback performance to stakeholders and guide changes. Manual data abstraction is onerous and not feasible. Reliance on receiving new reports from busy health system IT groups is challenging. Therefore, we sought to address this unmet need at our hospital by developing a novel ERAS Datamart system.

**Objective:** Our objectives were to develop a novel Datamart and Tableau dashboard to (1) enable continuous analysis of data, harvested directly from the electronic medical record (EMR), measure compliance and outcomes, and (2) enable end users (e.g., an ERAS coordinator) to create reports customized based on surgical procedure types, requested data variables, and custom date ranges.

**Methods:** After "buy-in" from hospital leadership and other stakeholders, data metrics were identified and categorized according to phase of care, that is, preoperative, intraoperative, and postoperative. A multidisciplinary team reviewed *International Classification of Diseases, Tenth Revision* procedure codes to capture EMR data for patients undergoing ERAS procedures. IT was given a master list with metric names, definitions, and screenshots of the discrete field in the EMR to assist with building the metrics. Validations of the novel Datamart were done against known ERAS patient populations maintained by the surgery clinic.

**Results:** The Datamart and Tableau dashboard has been built, is functional, and contains over 17,000 patients across 5 ERAS service lines: colorectal (n=1742), joint replacement (n=4235), surgical oncology (n=941), bariatric (n=1130), and cesarean section (n=9390). Currently, 56 metrics spanning the perioperative period have been validated across these populations. Reports can be tailored according to patients, time frames, and metrics. If desired, patient-level raw data can be exported for statistical analyses. Two use cases (total joint replacement and surgical oncology ERAS programs) are presented showing how the Datamart can be used.

**Conclusions:** Discrete fields within an EMR can be successfully captured into a novel Datamart and visualized using a custom Tableau dashboard for providing stakeholder feedback, facilitating quality improvement analyses, and auditing pathways.

***(JMIR Perioper Med 2026;9:e82472)*** doi:10.2196/82472

**KEYWORDS**

data monitoring; enhanced recovery; enhanced recovery after surgery; ERAS; perioperative outcomes; quality improvement; web platform

## Introduction

Enhanced recovery after surgery (ERAS) programs have transformed perioperative care by implementing evidence-based interventions that aim to standardize patient care and management, decrease resource utilization, expedite recovery, and improve patient outcomes [1-3]. The success and efficacy of ERAS programs are most likely achieved through the implementation of a comprehensive approach that bundles care for patients undergoing elective surgery, encompassing approximately 20 care elements [1,4].

It is increasingly recognized that as ERAS programs increase in size, it is very challenging to monitor and track bundle elements to feedback performance and guide outcomes. Traditional methods usually rely on manual data abstraction, frequent reports generated by hospital IT systems, or the use of third-party data warehouses. As ERAS programs grow in size and complexity, manual data abstraction becomes impractical due to time demands, error risk, and challenges with real-time analysis. Basic data points, such as length of stay, are easier to track, but capturing complex metrics, such as total opioid use (e.g., oral morphine equivalents), is often not feasible. Reliance on an IT report strategy is typically limited by very long delays in obtaining data reports from hospital IT workers who are usually burdened with many requests. The use of third-party data warehouses, for example, ERAS Interactive Audit System (EIAS), offers an alternative but raises concerns about data security, control, costs, system downtimes, and limited flexibility [5]. To address the above limitations, our institution created a novel dynamic Datamart dashboard.

## Methods

### Overview

Stony Brook University Hospital is a tertiary care academic medical center on Long Island, New York. Its first ERAS program (lumbar spine fusion) began in 2016, and an additional 9 ERAS programs were subsequently added. As the program grew, this revealed the unmet need for how to efficiently capture and monitor compliance and outcomes across a large number of patients.

As described in more detail below, the process for creating this system included: (1) leadership support and data governance, (2) validated identification of relevant patients to be included in the Datamart, (3) metric identification and validation, and (4) Tableau visualization as the user interface.

### Leadership Support and Data Governance

Under an institutional quality assurance program, a guideline was developed to map the creation of a novel Datamart and Tableau dashboard and govern the data extracted. To prioritize this effort, a value statement was presented to institutional leadership. This statement provided background information on the institution's ERAS programs, highlighted their prior success, and outlined the intended purpose of the Datamart and Tableau dashboard, such as monitoring and improving compliance, reducing errors associated with manual data abstraction, and limiting frequent report requests made to IT.

After approval, a Global ERAS Data Governance plan was conceived with policies and procedures for protecting and using ERAS data. These included how the data would be stored and protected, who would have access to data, and how data would be managed (e.g., requests for aggregate and patient-level reports, quality assurance (QA) analyses, and institutional review board–approved research projects). The Global ERAS Data Governance plan was subsequently signed by applicable departmental chairpersons to ensure data analysis was conducted in accordance with institutional standards and to protect against breaches of protected health information.

### Identification of Relevant Patients

#### *Choice of International Classification of Diseases, Tenth Revision Codes Methods*

We considered several possible strategies for identifying relevant patients for a given ERAS pathway. The hospital's operating room schedule, that is, planned surgical procedure, provides information on the "planned" procedure; however, it does not accurately reflect the "actual" surgery performed. Current procedural terminology professional billing codes were not used since the hospital's IT department did not have direct access to them. Therefore, we decided to use the *International Classification of Diseases, Tenth Revision* (*ICD-10*) Procedure Coding System (PCS) since this was feasible and these codes are believed to be accurate as they are used for hospital billing purposes.

To ensure the accurate identification of the ERAS patient population, the team collaborated closely with surgical leads from each ERAS pathway. Surgical procedures were sent to the coding department to identify the *ICD-10* PCS associated with specific procedures. The coding department supplied the leading 4 digits of all *ICD-10* PCS for the specified procedures. These digits encompass the section, body part, root operation, and where relevant, the body part of the given procedure. This preliminary list was forwarded to the IT department, which then appended the remaining digits of the *ICD-10* PCS, corresponding to the approach, device, and qualifier, for the procedures performed. Subsequently, a multidisciplinary team reviewed the complete *ICD-10* PCS to ensure the accurate capture of electronic medical record (EMR) data for patients undergoing ERAS procedures. This task required careful attention due to the complexities and overlaps within *ICD-10* PCS across multiple procedure types.

To ensure a comprehensive approach, a strategy was developed that included both *ICD-10* procedure codes and additional criteria, such as surgical case procedure name and other associated details (eg, associated *ICD-10* PCS, ambulatory surgical center vs main operating room, surgeon). This established a robust, multistep process for accurately pinpointing the desired patient population. This method ensured that only patients who received care associated with an ERAS pathway were included, enhancing the precision and reliability of the data captured for monitoring and analysis.

#### *Population Validation Methods*

Next, we validated the accuracy of using these *ICD-10* procedure codes to identify the desired patient population. This

validation aimed to identify instances where incorrect patients (non-ERAS patients) were erroneously included or, at the other extreme, ERAS patients were missing (i.e., not included). Admit type was then utilized to refine the patient population based on the urgency of admission, categorized as urgent, emergent, or elective. Since ERAS patients almost always fall into the "elective" category, this refinement allowed for an additional method for validating patient populations. Exception reports were generated and scheduled for automated reporting of patients who met the *ICD-10* PCS criteria but were admitted urgently or emergently. These records were then cross-referenced with the EMR to verify their eligibility for inclusion in the Datamart, ensuring only accurate ERAS populations were maintained. Several validations of the Datamart data were then performed by comparing the dataset against known ERAS patient populations maintained by the surgery clinic or stored in a REDCap (Research Electronic Data Capture) database.

### Metric Identification and Validation Methods

ERAS programs usually include many (e.g., ≥20) best practice elements. For example, the use of nonopioid analgesics, such as acetaminophen and nonsteroidal medications, is prioritized to minimize opioid use [6]. Since there are no national benchmarks for these programs to identify metrics of interest, key stakeholders (e.g.surgeons, anesthesiologists, hospital quality department personnel) were engaged. Metrics were selected, defined, and aligned with institutional interest, key performance, and patient safety indicators, and the American College of Surgeons National Surgical Quality Improvement Program (ACS NSQIP) by these stakeholders. We also sought to identify metrics that would be broadly applicable to most elective surgical procedures since the proposed Datamart system would be used for all ERAS programs.

Given the variations in documentation that can occur across an institution, it was crucial to investigate how and when an EMR field was completed. For example, nursing staff across different surgical specialties might record specifics related to urinary catheterization in separate discrete fields within the EMR. Additionally, there could be discrepancies in other charting practices as well, such as documenting ambulation as the "number of steps" taken versus "number of laps" taken. Metrics were validated with each ERAS population to identify discrepancies and ensure that the discrete field identified could be applied to the majority of surgical populations. This validation of metrics was carried out in multiple phases (e.g., 6 - 10 metrics at a time with each population) to alleviate the burden of mass validation. Metric validations were performed against known ERAS populations using manual chart review to ensure accurate data extraction into the Datamart from discrete EMR fields.

Once finalized, the metrics were categorized according to phase of care: preoperative, intraoperative, postoperative, and discharge. IT was provided with metric names, definitions, and screenshots of the discrete fields in the EMR to assist with building the metrics. To compartmentalize the data in the Datamart, these phases were bucketed into 7 categories, encompassing various areas of care. These categories included patient characteristics, preoperative care, operating room, postoperative fluid, postoperative multimodal analgesia, postoperative opioid, and postoperative patient experience.

### User Interface (Tableau Visualization) Methods

Tableau was chosen as the software for enabling a web-based dashboard user interface due to its on-demand filtering options and Health Insurance Portability and Accountability Act–compliant capabilities. An open query was established between our institution's EMR relational database and the Tableau dashboard. This setup allowed automated data extracts from the EMR to the Tableau dashboard according to the desired export rate. Multiple viewpoints and designs were trialed through various builds created by IT. Essential filtering options, such as specified time frames, surgical specialty, surgical procedures, and other specific metrics, were established and selected based on their relevance and informativeness for patients receiving care related to ERAS.

### Ethical Considerations

This project was conducted as part of an institutionally approved quality improvement/quality assurance initiative aimed at optimizing perioperative care processes and monitoring ERAS program performance. In accordance with our institution's policies on human participants protections, quality improvement/quality assurance activities that are designed solely for internal program evaluation are not considered human participants research.

## Results

Results are presented in the same sequence of events (1-4) as described in the Methods section.

### Leadership Support and Data Governance

An ERAS data governance guidance plan was established. The document outlined the appropriate use of the ERAS data warehouse, including storage, protection from data breach and leak of protected health information, access to data, and ensuring data analysis is in accordance with institutional standards. Data access is managed by a data access group. Requests for data must be submitted in writing to the data access group for review and approval. Data can be used for quality and research scholarly activities.

### Validation of Patient Selection Using *ICD-10* Procedures Codes

Initially, 50 metrics were trialed with *ICD-10* PCS for ERAS colorectal procedures. The patient population yielded from these codes returned many patients, more than 2-fold, compared to the known ERAS colorectal population. Preliminary validations revealed the need to (1) exclude surgery types such as emergent or urgent, (2) exclude overly broad or ambiguous *ICD 10* procedure codes, and (3) screen for potential inclusion of the planned surgical case procedure. After validating the colorectal population, the initial metrics were tested in the surgical oncology and joint replacement ERAS populations. We then backfilled data to 2015 to include pre-ERAS populations.

After final validations, the Datamart (last 6.5 y) contains more than 17,000 patients, consisting of colorectal (n=1742), surgical oncology (n=941), joint replacement (n=4235), bariatric (n=1130), and cesarian section (n=9390). In addition, the Datamart contains over 3000 archived cases from 2015 to 2018.

## Metric Identification and Validation

Over 100 metrics of potential interest, spanning the perioperative period and including patient characteristics, were identified. Several of these metrics represent the same metric measured at different time points over several days, such as peak pain on postoperative day (POD) 0, peak pain on POD 1, and peak pain on POD 2. The metrics were categorized according to the phase of care: preoperative, intraoperative, and postoperative. Currently, there are 56 metrics (demographics, surgical and anesthesia care, and postoperative endpoints) in the Datamart Textbox 1.

**Textbox 1.** Current metrics active in the Datamart.

1. Patient characteristics
   - Patient age, mean or median
   - Body mass index, mean or median
   - Diabetes diagnosis, #/%
   - Current smoker, #/%
   - Chronic opioid use (regular opioid use as listed on home medications), #/%
   - Associated diagnosis code, cancer, #/%
   - Associated diagnosis code, inflammatory bowel disease, #/%
   - Actual procedure completed, #/%
2. Preoperative care
   - Preprocedural bowel prep, #/%
   - Preprocedural oral antibiotics, #/%
   - Total functional status score (sum of activities of daily living score on the day of surgery in the preoperative area), mean or median
   - Hemoglobin $A_{1c}$ within 90 days (closest value prior to surgery start time), mean or median
   - Hemoglobin within 30 days (closest value prior to surgery start time), mean or median
3. Operating room
   - Laparoscopic procedure, #/%
   - Rectal procedure, #/%
   - Length of surgery, minutes, mean or median
   - Total intravenous (IV) fentanyl, mcg, mean or median
   - Received epidural or spinal, #/%
   - Intraoperative crystalloid IV fluids, mL (sum of normal saline, lactated ringers, and D5W), mean or median
4. Postoperative opioid
   - Day of surgery (DOS): oral morphine equivalent, mg, mean or median
   - Postoperative day (POD) 1: oral morphine equivalent, mg, mean or median
   - Postoperative day (POD) 2: oral morphine equivalent, mg, mean or median
5. Postoperative multimodal analgesia
   - Day of surgery
     - Oral acetaminophen, #/% of patients receiving any amount
     - IV acetaminophen, #/% of patients receiving any amount
     - IV or oral nonsteroidal anti-inflammatory drug (NSAID), #/% of patients receiving any amount
   - Postoperative day 1:
     - Oral acetaminophen, #/% of patients receiving any amount
     - IV acetaminophen, #/% of patients receiving any amount
     - IV or oral NSAID, #/% of patients receiving any amount
     - Multimodal agents, # of agents received, mean or median
   - Postoperative day 2:
     - Oral acetaminophen, #/% of patients receiving any amount
     - IV acetaminophen, #/% of patients receiving any amount
     - IV or oral NSAID, #/% of patients receiving any amount

- Multimodal agents, # of types of agents received, mean/median

- Postoperative day 3:
  - Oral acetaminophen, #/% of patients receiving any amount
  - IV acetaminophen, #/% of patients receiving any amount
- Total IV acetaminophen (DOS through POD 3), mg, mean or median
- Total oral acetaminophen (DOS through POD 3), mg, mean or median

6. Postoperative fluid
   - Day of surgery
     - Net input and output, mL, mean or median
     - Total input, mL, mean or median
   - · Postoperative day 1:
     - Net input and output, mL, mean or median
     - Total input, mL, mean or median

7. Postoperative patient experience
   - Postoperative patient-controlled analgesia use, hours, mean or median
   - Peak pain score on day of surgery, mean or median
   - Peak pain score on postoperative day 1, mean or median
   - Peak pain score on postoperative day 2, mean or median
   - Postoperative urinary straight catheterization on day of surgery, #/%
   - Postoperative urinary straight catheterization on postoperative day 1, #/%·
   - Postoperative insertion of urinary catheter (insertion occurs more than 2 hours after surgery stop or removal in operating room), #/%
   - Postoperative duration of urinary catheter, hours, mean or median
   - Time to first flatus or bowel movement, hours, mean or median
   - Postoperative duration of nasogastric tube, hours, mean or median
   - Postoperative insertion of nasogastric tube (insertion occurs more than 2 hours after surgery stop or removal in operating room), #/%
   - Delta creatinine (peak value within 72 hours postoperatively minus preoperative value), mean or median
   - Postoperative length of stay, days, mean or median
   - Discharge opioids (oral morphine equivalents) prescribed, mg, mean or median

Unless specified otherwise, units of measurement for continuous variables are mg and categorical end points are yes/no.

## Tableau Visualization

### *Overview*

The Datamart dashboard was designed to incorporate a user-friendly interface, allowing users to easily generate customizable reports. It offers tools to filter data by patient population, time frames, and specific metrics, enabling a focused analysis of ERAS outcomes (e.g., compare before and after implementation outcomes, track changes in protocols over specific time frames).

The user is first presented with a screening interface where patient population, time frames, and age parameters can be defined (Figure 1). By hovering over the colored square, the user can also review certain patient and surgery-specific information, such as associated *ICD-10* PCS, surgeon, and date of surgery (Figure 2). Patients can be deselected from the population if, for example, erroneous classification of an emergency trauma patient as an elective case.

**Figure 1.** User interface (screening). User interface depicting screening options. The user can select time frames, patient age range, and surgical population. The user is presented with the patient name, arrival and discharge date and time, patient age on day of surgery, qualifier for inclusion in the Datamart (*International Classification of Diseases, Tenth Revision* [*ICD-10*] *Procedure Coding System* [PCS] and/or surgical case procedure), and qualifier description. Patients can be selected or deselected on the right. Hovering over the colored square yields additional screening data (see Figure 2).

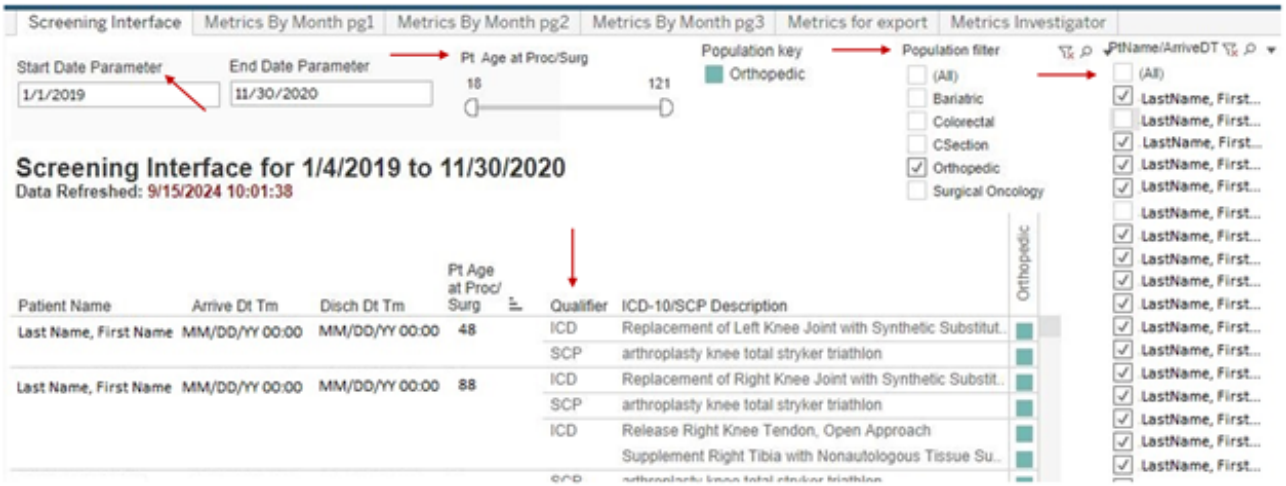
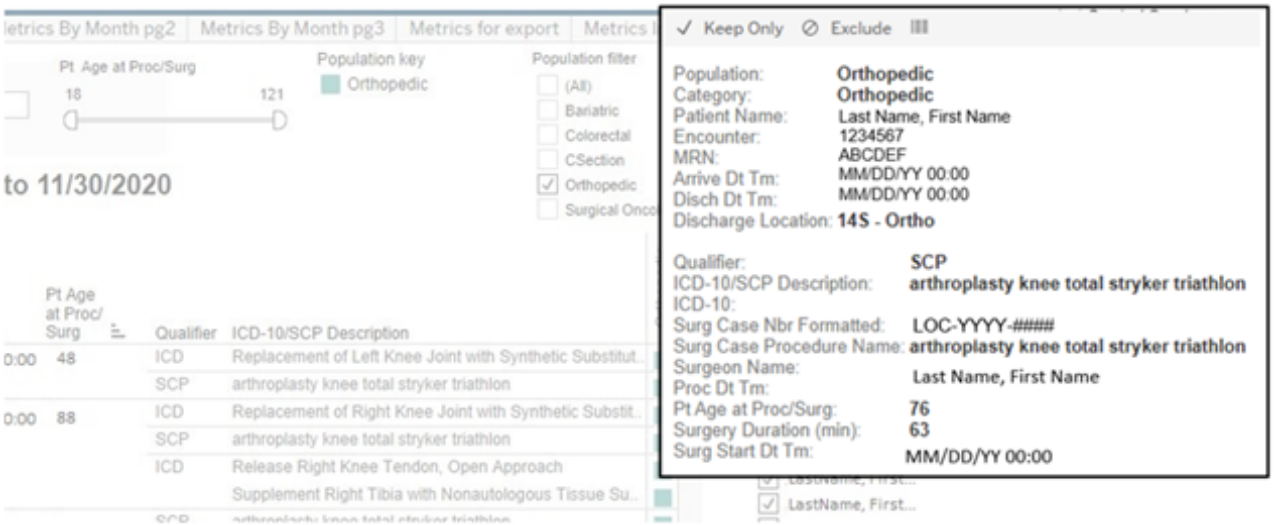
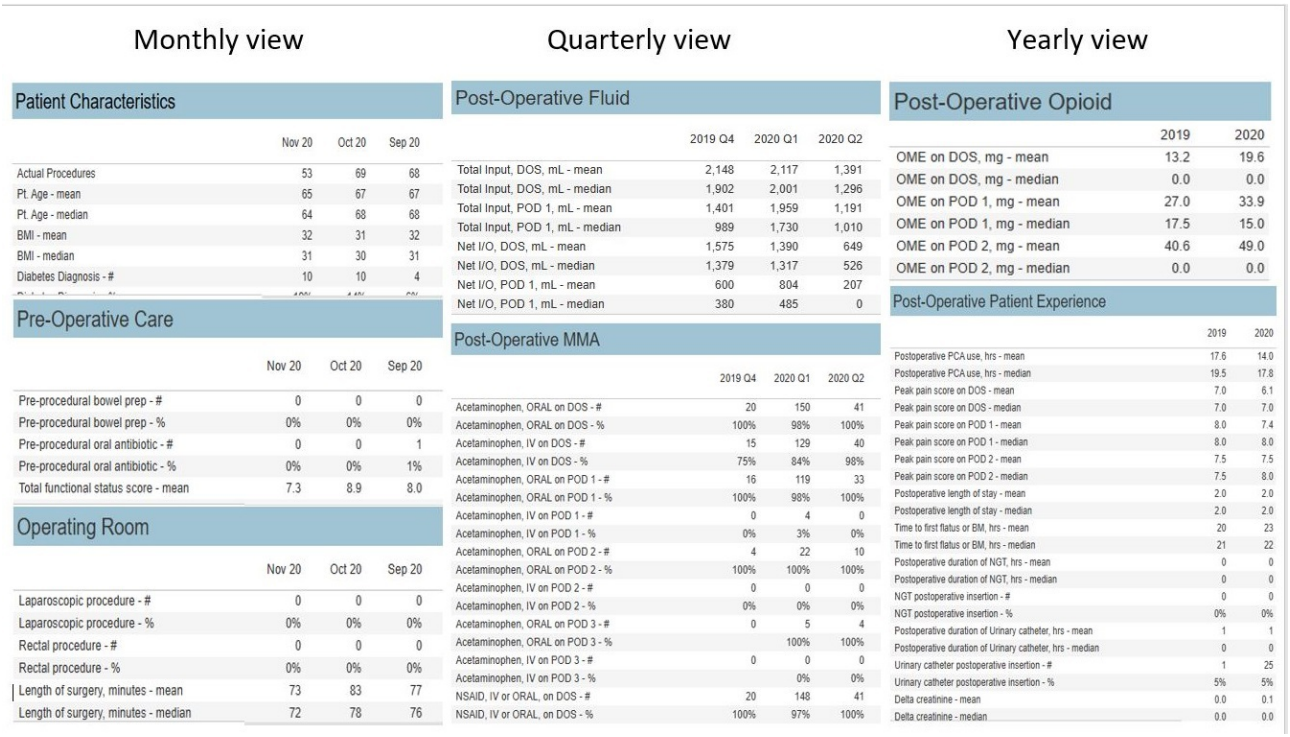



**Figure 2.** User interface (related screening). Additional screening information is available to the user. Hovering over the colored square next to the *International Classification of Diseases, Tenth Revision* (*ICD 10*)/surgical case procedure (SCP) description "pops out" information related to the selected case. This includes discharge location, surgical case number, surgeon name, and length of surgery.




On subsequent screens, the user can determine the metrics for analysis and how to view the data (Figure 3). Filters can be customized to focus on specific metrics or patient outcomes. For example, to focus on total acetaminophen use, one can view more details, such as "total IV acetaminophen on post-op day 1 (mean or median values)." Individual elements, such as opioid administration or fluid management details, can also be isolated for deeper analysis. Additionally, data can be displayed in weekly, monthly, quarterly, or yearly summaries and saved and exported as tables.

**Figure 3.** User interface (metrics). After procedures have been selected for review, the user can determine the metrics for analysis. Metrics can be viewed weekly, monthly, quarterly, or annually. Individual elements, such as opioid administration or fluid management details, can also be isolated for deeper analysis. DOS: day of surgery; IV: intravenous;nasogastric tube; MMA: multimodal analgesia; NGT: NSAID: nonsteroidal anti-inflammatory drug; OME: oral morphine equivalent; PCA: patient-controlled analgesia; POD: postoperative day.




For further in-depth analyses, for example, inferential statistical analysis, users can export patient-level raw data in either comma-separated value or Excel format. These data can be exported in a deidentified manner, safeguarding against the unintentional disclosure of protected health information when exiting the secure Datamart dashboard. By using these features, the Datamart permits users to analyze ERAS metrics with precision, adapt to different populations, and investigate trends and outcomes.

### *Use Case Examples: ERAS Total Joint Replacement and Surgical Oncology*

Using total joint replacement as a use case example, we sought to compare patient care prior to the initiation of ERAS (pre-ERAS n=693 cases) and 12 months post-ERAS implementation (n=563 cases). Figures 4 and 5 show the impact of an ERAS program on pain using the data harvested from the Datamart, notably an apparent improvement in several pain-related metrics. Of note, the Tableau dashboard does not calculate or show error bars for continuous variables; the calculation of these requires export of the raw data for statistical analysis. Therefore, we opted to present mean and median, since if they are similar, it suggests that the data are normally distributed and the mean value is not inflated due to a few large outliers. Therefore, for most day-to-day QA purposes, the presentation of mean and median is sufficient, but for more rigorous quantitative analysis, for example, inferential statistics for hypothesis testing, the export of patient-level raw data allows that capability.

**Figure 4.** Enhanced recovery after surgery (ERAS) total joint replacement program: impact on pain. The Tableau dashboard does not display error bars for continuous variables; these require the export of patient-level raw data for statistical analysis, which is a capability of the Datamart. As ERAS pathways are primarily a quality assurance (QA) initiative, we report mean and median values, as their similarity suggests a roughly normal distribution without major outliers. DOS: day of surgery; IV: intravenous; NSAID: nonsteroidal anti-inflammatory drug; OME: oral morphine equivalent; POD: postoperative day.

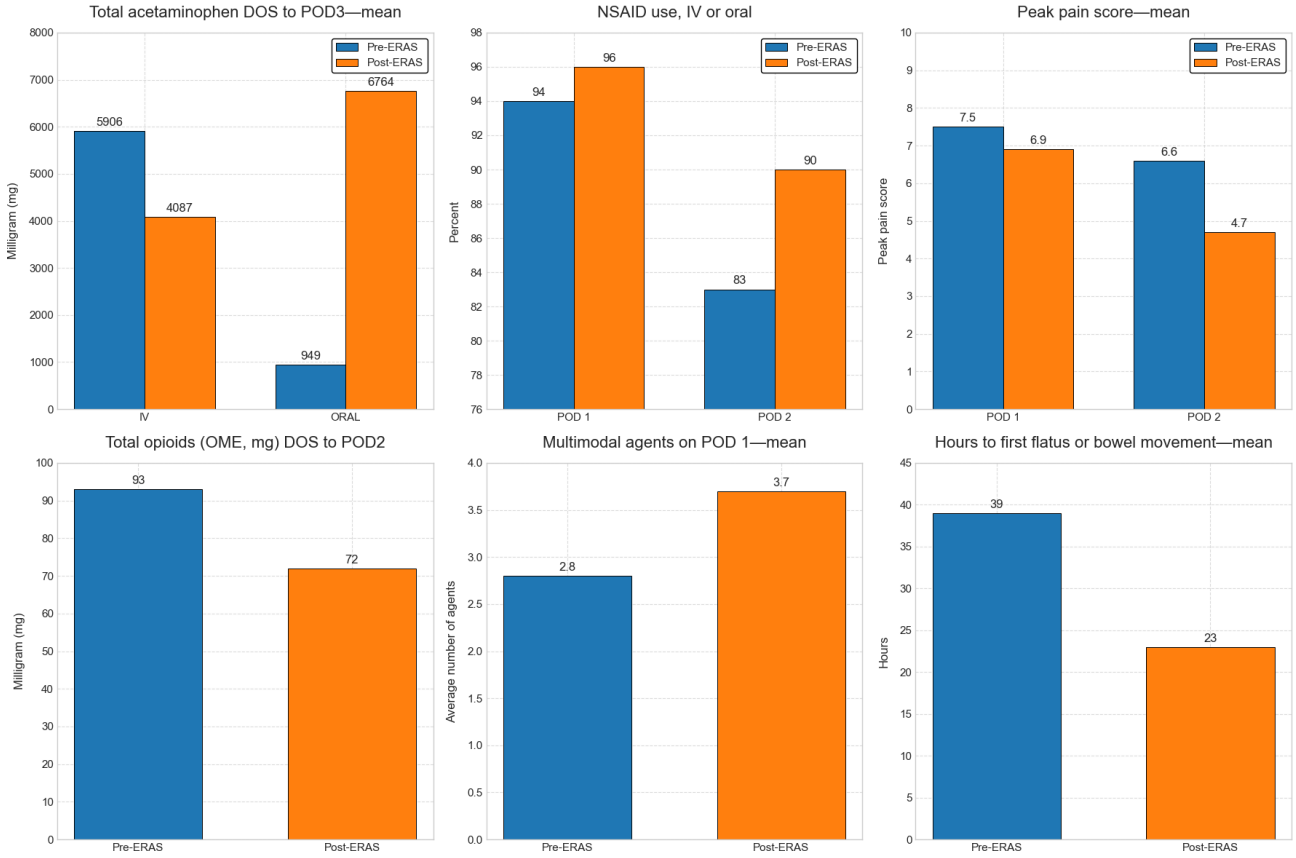
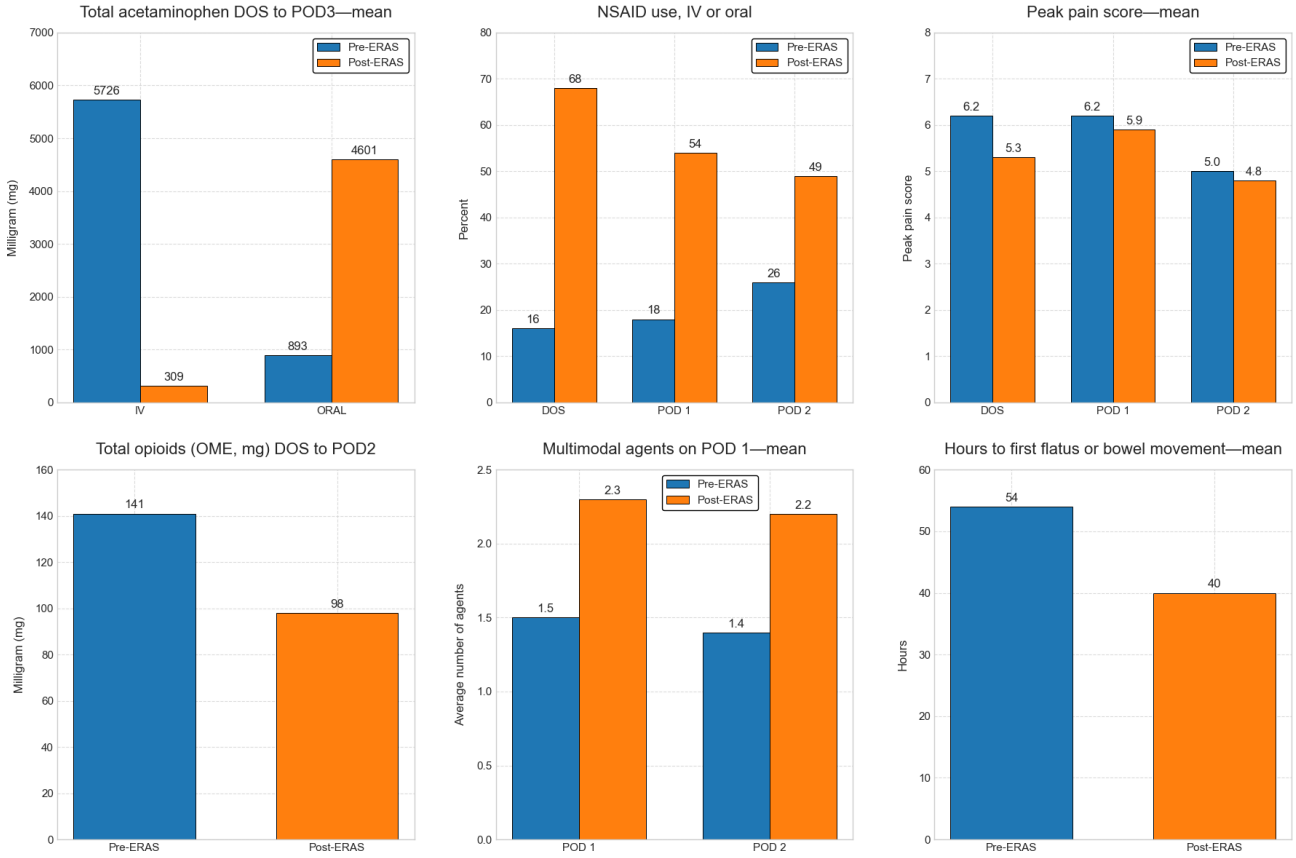


**Figure 5.** Enhanced recovery after surgery (ERAS) surgical oncology program: impact on pain. The Tableau dashboard does not display error bars for continuous variables; these require export of patient-level raw data for statistical analysis, which is a capability of the Datamart. As ERAS pathways are primarily a quality assurance (QA) initiative, we report mean and median values, as their similarity suggests a roughly normal distribution without major outliers. DOS: day of surgery; IV: intravenous; NSAID: nonsteroidal anti-inflammatory drug; OME: oral morphine equivalent; POD: postoperative day.




We noticed similar findings for our Surgical Oncology ERAS program (Figure 5).

## Discussion

### Principal Results

The novel Datamart dashboard has been a transformative tool at our hospital by enabling continuous analysis of ERAS programs and patient outcomes. By automating data extraction from the EMR, the Datamart eliminates the need for manual data abstraction, likely reducing errors and improving data accuracy. This centralized system currently serves as the primary data resource for our institutional ERAS programs, streamlining the monitoring of bundle elements and outcomes at both the individual and aggregate levels. Successful adoption required early stakeholder engagement, quarterly dashboard reviews, and review at quarterly stakeholder meetings. Barriers included initial unfamiliarity with Tableau capabilities and competing clinical priorities. These were mitigated through targeted training and dashboard use in routine quality meetings.

Key features of the Datamart include categorized metrics across perioperative phases, tailored reporting capabilities, and the ability to export data. The level of granularity of the Datamart provides actionable insights into ERAS pathway efficacy, supporting evidence-based decision-making.

### Limitations

Despite its many advantages, the Datamart has some limitations that warrant mention. As with any EMR-based form of data capture, it relies on accurate charting by clinicians, which is not infallible. For example, postoperative ambulation was not included in this iteration, as documentation practices are inconsistent. Moreover, patient-reported outcomes, such as satisfaction and functional recovery, are not currently documented in the EMR, which prevents their inclusion in the Datamart.

Accurately identifying patients on an ERAS pathway is another challenge. The Datamart relies on *ICD-10* PCS for surgical procedure identification, which can lead to the inadvertent inclusion of non-ERAS cases. For example, procedures such as hemorrhoidectomy (non-ERAS) and hemicolectomy (ERAS) share the same *ICD-10* PCS, requiring manual filtering to exclude nonrelevant cases. While procedures, such as total knee replacement or cesarean section, are easier to identify due to constrained coding options, consistent oversight is critical for accurate data classification.

An intentional 2-month lag in data import further limits real-time analysis. However, this delay ensures the accuracy of *ICD-10* coding and surgical procedure inclusion, contributing to the integrity of the dataset.

## Comparison With Prior Work

Health care auditing has significantly evolved with the advent of digital dashboards, data warehouses, and interactive audit systems. Before widespread dashboard integration, many institutions relied on manual data abstraction and semiautomated systems. Manual abstraction is cumbersome and relies on human resources to extract data from the EMR to input into a repository for analysis (e.g., REDCap, Microsoft Excel) [7-9]. Semiautomated audits (e.g. IT-generated reports), which at our institution can take 9 months or longer, add to the resource burden. Some institutions combine EMR data extraction with administrative databases (e.g., NSQIP) to build a centralized reporting structure [2,10]. However, reliance on these reports poses challenges such as delayed access to real-time data, lack of customization, and risk of system downtimes.

For some institutions with ERAS pathways, third-party data warehouses (e.g., EIAS) may be an option [5]. While they offer additional storage and analysis options, third-party warehouses present issues of data security, higher costs, and limited control over information (see Table 1). In contrast, our in-house customizable Datamart can support diverse ERAS populations and continuous improvements through iterative refinements based on user feedback and technological advancements.

**Table .** Comparison between the Datamart and third-party options.

| Feature | Datamart+Tableau | EIAS[a] | ACS[b] NSQIP[c] |
|---|---|---|---|
| Data control | Full institutional control over data | Vendor-controlled; limited flexibility | External benchmarking database |
| Customization | Highly customizable by procedure type, metrics, and time frame | Some customization possible, but limited | Standardized measures; less customizable |
| Real-time access | Near real-time (2-month lag for data integrity) | Typically delayed, relies on upload | Reports released quarterly or semiannually |
| Cost | Internal development; no licensing fees | Requires subscription or license fees | Expensive participation and data access |
| Data security | Remains within institutional firewall; HIPAA[d]-compliant | Data housed externally; possible security concerns | Data deidentified but externally stored |
| Metric flexibility | Fully institution-defined (≥56 metrics currently used) | Limited to ERAS[e]-recommended fields | Fixed set of standard metrics |
| Scalability | Easily scalable across services and use cases | Limited by system design and vendor | Only covers specific surgeries (e.g., colectomy) |

[a]EIAS: ERAS Interactive Audit System.

[b]ACS: American College of Surgeons.

[c]NSQIP: National Surgical Quality Improvement Program.

[d]HIPAA: Health Insurance Portability and Accountability Act.

[e]ERAS: enhanced recovery after surgery.

While we were unable to find reports in the literature of the use of a datamart with dashboard visualization tools to support ERAS programs, these tools have been described in other types of quality, clinical, and research programs. Institutionally developed data infrastructures have been used to aggregate disparate data sources into unified repositories to support research and clinical audits [11-13]. Interactive dashboards (e.g., Tableau and Qlik) have been implemented to visualize clinical performance and facilitate quality improvement efforts [14-16]. Thus, enabling data-driven feedback to stakeholders. The Datamart combines the filtering and query capabilities of a unified data repository with the visualization tools of interactive dashboards to improve patient outcomes, enhance adherence to protocols, improve interdisciplinary communication, and support decision-making.

## Conclusions

The Datamart represents a significant advancement in ERAS program management. Unlike third-party systems, such as the EIAS or ACS NSQIP, the Datamart provides full institutional control over data, customizable metrics aligned with local priorities, and flexible reporting capabilities. Although third-party systems can be limited by fixed datasets, external data hosting, and reporting delays, the Datamart enables near real-time internal data access, tailored QA tracking, and the ability to refine metrics, as clinical needs evolve. This offers a robust, centralized, and automated approach to data monitoring and analysis without recurring licensing costs or dependence on vendor timelines. The ability to reduce manual abstraction, improve data accuracy, and evaluate intervention makes the Datamart a vital tool for enhancing perioperative care.

There are several promising directions for the enhancement and expansion of this system. Integrating predictive analytics, automated alerts, natural language processing, and machine learning into the Datamart could enable users to better anticipate complications and tailor interventions [11,12]. For example, predictive models could identify patients at high risk for ERAS noncompliance or adverse outcomes, allowing proactive adjustments to care plans. Automated alerts embedded within the dashboard could notify clinicians in real time when critical metrics fall outside expected ranges, supporting timely interventions and reducing preventable complications.

Expanding the dashboard to encompass additional surgical specialties and incorporating patient-reported outcomes, through structured EMR fields or digital surveys, into the Datamart could further enhance applicability and patient-centered care. These innovations would transform the Datamart from a retrospective monitoring tool into a dynamic, decision-support platform that drives continuous improvement in perioperative care.

As ERAS programs grow in scope and complexity, the need for scalable, adaptable solutions to implementing and monitoring evidence-based care and patient outcomes is increasingly evident. This novel Datamart dashboard addresses many of these challenges while providing a foundation for ongoing innovation. Although developed within Cerner, the Datamart framework is adaptable to other EMRs, provided discrete data fields are available. Implementation requires collaboration with nursing, anesthesiology and surgical services, institutional IT, along with Tableau or similar visualization tools. Institutions seeking to improve ERAS monitoring may consider adapting the framework described here, tailoring metric section and dashboard design to their local EMR environment and clinical priorities.

## Data Availability

The datasets used are not available.

## Authors' Contributions

Conceptualization: EB-G (lead), SMS (equal)
Data curation: EB
Formal analysis: AS (supporting), EB (lead), JN (supporting), SMS (supporting)
Funding acquisition: EB-G
Investigation: EB (equal), SMS (equal)
Methodology: EB (equal), EB-G (lead), SMS (equal)
Project administration: EB (supporting), EB-G (lead), SMS (equal)
Resources: EB
Software: EB
Supervision: EB-G
Validation: EB (equal), SMS (equal)
Visualization: EB-G (lead), SMS (supporting)
Writing – original draft: EB-G (equal), SO (equal), SMS (equal)
Writing – review & editing: AS (supporting), EB (supporting), EB-G (lead), JN (supporting), SMS (equal), SO (equal)

## Conflicts of Interest

None declared.

## References

1. Gramlich LM, Sheppard CE, Wasylak T, et al. Implementation of enhanced recovery after surgery: a strategy to transform surgical care across a health system. Implement Sci 2017 May 19;12(1):67. [doi: 10.1186/s13012-017-0597-5] [Medline: 28526041]
2. Thiele RH, Rea KM, Turrentine FE, et al. Standardization of care: impact of an enhanced recovery protocol on length of stay, complications, and direct costs after colorectal surgery. J Am Coll Surg 2015 Apr;220(4):430-443. [doi: 10.1016/j.jamcollsurg.2014.12.042] [Medline: 25797725]
3. Singh SM, Liverpool A, Romeiser JL, Thacker J, Gan TJ, Bennett-Guerrero E. Types of surgical patients enrolled in enhanced recovery after surgery (ERAS) programs in the USA. Perioper Med (Lond) 2021 Apr 27;10(1):12. [doi: 10.1186/s13741-021-00185-5] [Medline: 33902705]
4. Carey S, Hogan S. Failure in systems and culture: barriers that prevent implementation of evidence-based fasting times for patients in the acute care setting. JPEN J Parenter Enteral Nutr 2021 Jul;45(5):933-940. [doi: 10.1002/jpen.1961] [Medline: 32654214]
5. ERAS interactive audit system - EIAS by Encare. Encare Enhancing Eras. URL: https://encare.net/products-and-services/eras-interactive-audit-system-eias/ [accessed 2024-12-12]
6. Langnas EM, Matthay ZA, Lin A, et al. Enhanced recovery after surgery protocol and postoperative opioid prescribing for cesarean delivery: an interrupted time series analysis. Perioper Med (Lond) 2021 Nov 15;10(1):38. [doi: 10.1186/s13741-021-00209-0] [Medline: 34775985]
7. Carter J. Fast-track surgery in gynaecology and gynaecologic oncology: a review of a rolling clinical audit. ISRN Surg 2012;2012:368014. [doi: 10.5402/2012/368014] [Medline: 23320193]
8. Archibald LH, Ott MJ, Gale CM, Zhang J, Peters MS, Stroud GK. Enhanced recovery after colon surgery in a community hospital system. Dis Colon Rectum 2011 Jul;54(7):840-845. [doi: 10.1007/DCR.0b013e31821645bd] [Medline: 21654251]

9. Sinha R, Verma N, Bana R, Kalidindi N, Sampurna S, Mohanty GS. Intra- and post-operative outcomes in benign gynaecologic surgeries before and after the implementation of enhanced recovery after surgery protocols: a comparison. J Minim Access Surg 2023;19(1):112-119. [doi: 10.4103/jmas.jmas_42_22] [Medline: 36722536]
10. Ripollés-Melchor J, Abad-Motos A, Díez-Remesal Y, et al. Association between use of enhanced recovery after surgery protocol and postoperative complications in total hip and knee arthroplasty in the Postoperative Outcomes Within Enhanced Recovery After Surgery Protocol in Elective Total Hip and Knee Arthroplasty Study (POWER2). JAMA Surg 2020 Apr 1;155(4):e196024. [doi: 10.1001/jamasurg.2019.6024] [Medline: 32049352]
11. Marazzi F, Tagliaferri L, Masiello V, et al. GENERATOR Breast DataMart—the novel breast cancer data discovery system for research and monitoring: preliminary results and future perspectives. J Pers Med 2021 Jan 22;11(2):65. [doi: 10.3390/jpm11020065] [Medline: 33498985]
12. D'Amario D, Laborante R, Delvinioti A, et al. GENERATOR HEART FAILURE DataMart: an integrated framework for heart failure research. Front Cardiovasc Med 2023;10:1104699. [doi: 10.3389/fcvm.2023.1104699] [Medline: 37034335]
13. Hurst JH, Liu Y, Maxson PJ, Permar SR, Boulware LE, Goldstein BA. Development of an electronic health records datamart to support clinical and population health research. J Clin Transl Sci 2020 Jun 23;5(1):e13. [doi: 10.1017/cts.2020.499] [Medline: 33948239]
14. Lechner C, Rumpler M, Dorley MC, Li Y, Ingram A, Fryman H. Developing an online dashboard to visualize performance data—Tennessee newborn screening experience. Int J Neonatal Screen 2022 Sep 2;8(3):49. [doi: 10.3390/ijns8030049] [Medline: 36135349]
15. Dang QCL, Román E, Donner K, et al. Implementation and impact of a surgical dashboard on pediatric tonsillectomy outcomes: a quality improvement study. Laryngoscope Investig Otolaryngol 2024 Oct;9(5):e1315. [doi: 10.1002/lio2.1315] [Medline: 39281202]
16. Alhmoud B, Melley D, Khan N, Bonicci T, Patel R, Banerjee A. Evaluating a novel, integrative dashboard for health professionals' performance in managing deteriorating patients: a quality improvement project. BMJ Open Qual 2022 Dec;11(4):e002033. [doi: 10.1136/bmjoq-2022-002033] [Medline: 36588306]


## Abbreviations

**ACS:** American College of Surgeons
**EIAS:** ERAS Interactive Audit System
**EMR:** electronic medical record
**ERAS:** Enhanced Recovery after Surgery
***ICD-10*:** *International Classification of Diseases, Tenth Revision*
**NSQIP:** National Surgical Quality Improvement Program
**PCS:** Procedure Coding System
**POD:** postoperative day
**QA:** quality assurance
**REDCap:** Research Electronic Data Capture


*Edited by N Rohatgi; submitted 15.Aug.2025; peer-reviewed by AR Cochran, C McLane; revised version received 15.Dec.2025; accepted 18.Dec.2025; published 11.Feb.2026.*

*Please cite as:*
*Singh SM, Oster S, Bolze E, Sasson A, Nicholson J, Bennett-Guerrero E*
*A Novel Customizable Datamart and Tableau Dashboard to Monitor Multiple Enhanced Recovery After Surgery Programs: Development and Validation Study*
*JMIR Perioper Med 2026;9:e82472*
*URL: https://periop.jmir.org/2026/1/e82472*
*doi:10.2196/82472*

# Comparison Between Ultrasound and Magnetic Resonance Imaging Measurements of the Optic Nerve Sheath Diameter in Patients Undergoing Intracranial Surgery: Prospective Observational Single-Center Study

Mauricio Giraldo[1*], MD; Luz Maria Lopera[2*], MD; Aly Balbaa[3*], MD; Nelson Gonzalez[1*], MD; Raffael Pereira Cezar Zamper[1*], MD; Michael Mayich[4*], MD; Mel Boulton[5*], MD

[1]Department of Anesthesia and Perioperative Medicine, Schulich School of Medicine & Dentistry, Western University, London Health Sciences Centre, 339 Windermere Road, London, ON, Canada

[2]Department of Anesthesia and Perioperative Medicine, Faculty of Medicine, Dalhousie University, Queen Elizabeth II Health Sciences Centre, Halifax, NS, Canada

[3]Department of Family Medicine, Faculty of Medicine, McMaster University, Hamilton Health Sciences, Hamilton, ON, Canada

[4]Department of Neuroradiology, Schulich School of Medicine & Dentistry, Western University, London Health Sciences Centre, London, ON, Canada

[5]Department of Neurosurgery, Schulich School of Medicine & Dentistry, Western University, London Health Sciences Centre, London, ON, Canada

[*]all authors contributed equally

**Corresponding Author:**
Mauricio Giraldo, MD
Department of Anesthesia and Perioperative Medicine, Schulich School of Medicine & Dentistry, Western University, London Health Sciences Centre, 339 Windermere Road, London, ON, Canada

## Abstract

**Background:** Measuring the optic nerve sheath diameter (ONSD) with ultrasound is a promising, noninvasive way to estimate intracranial pressure (ICP). While magnetic resonance imaging (MRI) provides high-resolution imaging, it is less accessible in urgent or perioperative settings. Comparing ONSD measurements between ultrasound and MRI may help confirm the use of ultrasound in neurosurgical patients.

**Objective:** The aim of this study is to evaluate how closely ultrasound and MRI measurements of ONSD align in patients undergoing surgery for supratentorial brain tumors.

**Methods:** This prospective, single-center observational study included 50 adult patients scheduled for elective supratentorial tumor resection. ONSD was measured preoperatively using both transorbital ultrasound and MRI. Measurements were compared using Pearson and Spearman correlation coefficients, the intraclass correlation coefficient, and Bland-Altman analysis.

**Results:** The average ONSD measured by ultrasound was 5.94 (0.99) mm, compared to 5.75 (SD 1.08) mm via MRI. The two methods showed a strong correlation (Pearson $r$=0.88, $P$<.001) and good agreement (intraclass correlation coefficient=0.86). Bland-Altman analysis showed a mean bias of 0.19 mm (95% limits of agreement: –0.62 to 1.00 mm).

**Conclusions:** Ultrasound-based ONSD measurements closely matched those obtained by MRI in this patient group. These findings support the use of ultrasound as a practical tool for noninvasive ICP assessment in the perioperative care of patients with intracranial tumors.

***(JMIR Perioper Med 2026;9:e67480)*** doi:10.2196/67480

**KEYWORDS**

optic nerve; optic nerve diameter; intracranial pressure; ocular ultrasound; brain tumor

## Introduction

Accurate assessment of intracranial pressure (ICP) is essential in the management of patients with a range of neurological conditions [1,2]. Several tools are available for this purpose, including invasive monitoring devices, magnetic resonance imaging (MRI), computed tomography (CT), and ultrasonography (US) [3,4]. While invasive methods remain the gold standard, they carry risks and are not always feasible. MRI and CT offer practical alternatives but can be limited by cost, access, and the need for patient transport, particularly in acute or perioperative settings.

In recent years, ultrasonographic measurement of the optic nerve sheath diameter (ONSD) has gained attention as a practical,

noninvasive method for estimating ICP. Ultrasound is widely available at the bedside, relatively inexpensive, and quick to perform. A growing body of evidence supports its use, with studies showing good correlation between ONSD measured by ultrasound and both invasive ICP measurements and imaging-based assessments [5]. This technique has been validated in adult and pediatric populations, including neonates, and across a range of pathologies such as trauma, hydrocephalus, ischemic stroke, and brain tumors. Systematic reviews have also contributed to the evidence base [6,7]. For example, a 2011 meta-analysis by Moretti et al [8] demonstrated the diagnostic accuracy of ONSD US in detecting raised ICP. A 2018 review further emphasized its value, particularly when invasive methods are unavailable or contraindicated [9].

Despite its clinical promise, some uncertainties remain. One ongoing discussion involves the optimal cutoff for diagnosing elevated ICP. While there is no universally accepted threshold, ONSD values of 0.48 - 0.50 cm (4.8 - 5.0 mm) are commonly used to indicate ICP above 20 cm $H_2O$ or 20 mm Hg [10-15], with several studies reporting high sensitivity and specificity with these values. For instance, Kimberly et al (2008) [16] reported that an ONSD exceeding 5.0 mm was a strong predictor of raised ICP.

Technique is another area where standardization is still evolving. The typical approach involves measuring the sheath in the transverse plane, 5 mm posterior to the globe [17]. However, anatomical differences between individuals, particularly in the transverse diameter of the eyeball (ETD), may influence interpretation. Some researchers have proposed using the ONSD/ETD ratio to adjust for these variations, with values above 0.19 potentially indicating elevated ICP. While ETD ranges from 21 to 27 mm in healthy individuals, its use in routine clinical decision-making remains under investigation [17]. Recent publications also highlight the growing role of ocular ultrasound in ICP assessment [18,19]. One open-access study noted that bedside ONSD measurement can provide quick, clinically meaningful information during the perioperative period, especially when formal imaging is not immediately available. Their work supports the idea that ultrasound is becoming an increasingly practical addition to routine neurologic monitoring, helping clinicians recognize changes in intracranial dynamics without interrupting patient flow [19].

This study aims to explore the use of ultrasound-based ONSD measurement in a specific neurosurgical population: patients with supratentorial brain tumors. These patients are at risk for increased ICP due to mass effect, edema, or obstruction of cerebrospinal fluid pathways. While previous research has shown a strong correlation between ONSD measured by ultrasound and other modalities, our goal is to assess the agreement between ultrasound- and MRI-derived ONSD values obtained on the same day in a controlled preoperative setting. We also examine the feasibility and accuracy of ultrasound when performed by trained anesthesiologists, a group often involved in perioperative decision-making. Our findings may help clarify the role of ONSD US in situations where access to MRI or CT is limited or delayed.

We conducted a prospective observational study of 50 adult patients undergoing elective supratentorial brain tumor resections at London Health Sciences Center (University Hospital) between July 2021 and May 2023.

## Methods

### Patient Selection

Inclusion criteria were adults (≥18 y) scheduled for elective intracranial surgery involving supratentorial tumors. Patients with previous intraocular lens implantation were included in the study. Patients were excluded if they experienced hemodynamic instability, required emergent or repeat intracranial surgery, or had a history of ocular pathology, including ocular infection, ocular trauma, or prior ocular surgery. All participants provided written informed consent prior to enrollment.

### Ultrasound Examination

On the day of surgery, patients underwent bedside ocular ultrasound prior to anesthetic induction. Ultrasound was performed by a cardiac anesthesiologist formally trained in transthoracic and transesophageal echocardiography and point-of-care ultrasound (POCUS), who had also received a dedicated 2-hour session in standardized ocular ultrasound technique. The session followed current consensus recommendations and included both didactic and hands-on instruction.

Patients were positioned supine with the neck in a neutral position. A high-frequency linear probe (4 - 12 MHz, L12-4; Koninklijke Philips NV) was used with sterile water-soluble gel (Aquasonic 100, Parker Laboratories Inc) applied over the closed eyelid, following standard radiologic convention. The ultrasound marker was oriented toward the patient's right. The optic nerve was visualized in the transverse plane (Figure S1 in Multimedia Appendix 1), and measurements of the ONSD were taken 5 mm posterior to the retina (Figure S2 in Multimedia Appendix 2), perpendicular to the optic nerve axis. Each eye was scanned separately.

Three ONSD measurements were obtained per eye in the transverse plane, and the average of these was used for analysis (Figure S3 in Multimedia Appendix 3). If image quality was suboptimal, additional attempts were made until three acceptable measurements were recorded. All images were stored using a Philips Sparq ultrasound system and Q-Path 5.2.434 software (Telexy Corporation) and anonymized using numeric identifiers per institutional privacy protocols. The ultrasound results were not disclosed to the neuroradiologist performing the MRI measurements.

### MRI Assessment

All patients underwent preoperative brain MRI as part of their standard neurosurgical evaluation. MRI images were obtained using a Siemens Magnetom Vida 3T scanner (Siemens Healthineers). ONSD measurements were performed by a neuroradiologist using standard institutional technique in the axial plane, approximately 5 mm posterior to the globe (Figure S4 in Multimedia Appendix 4). The measurement included the

whole optic nerve sheath complex and was obtained at the widest diameter perpendicular to the nerve axis. The MRI results were not disclosed to the anesthesiologist.

### Data Collection

Demographic and clinical variables collected included age, sex, height, weight, and BMI, along with ONSD measurements from both ultrasound and MRI. These covariates were included based on prior literature suggesting that factors such as sex, age, and anthropometric measurements may influence baseline ONSD values or their interpretation in the context of raised ICP. Additionally, clinical characteristics such as tumor size, tumor location, and duration of symptoms were recorded, as these may influence ICP and, by extension, ONSD.

### Statistical Analysis

Data were analyzed using descriptive and inferential statistics. Continuous variables were summarized as means and standard deviations for normally distributed data or medians with interquartile ranges for nonnormal data, while categorical variables were summarized using counts and percentages. Normality of numerical variables was assessed using the Shapiro-Wilk test.

The relationship between ONSD measurements obtained by ultrasound and MRI was evaluated using the Pearson correlation coefficient for normally distributed variables and the Spearman rank correlation when normality assumptions were not met. Differences between ultrasound and MRI ONSD measurements were compared using paired *t* tests for normally distributed data or Wilcoxon signed-rank tests otherwise. Agreement between the two measurement methods was further assessed using the intraclass correlation coefficient (ICC) with a 1-way random-effects model, with significance tested using the *F* test.

A *P* value <.01 was considered statistically significant to account for multiple comparisons and reduce the risk of type I error in this small, exploratory sample. This threshold was defined a priori.

### Ethical Considerations

This study was reviewed and approved by the Research Ethics Board of Western University (protocol number 115045). The study was conducted in accordance with the ethical principles outlined in the Declaration of Helsinki and applicable institutional and national research guidelines. Written informed consent was obtained from all participants prior to enrollment. No compensation was provided to participants.

## Results

### Patient Demographics

A total of 50 patients undergoing brain tumor resection were included in the study, comprising 25 males and 25 females. Participant ages ranged from 21 to 80 years, with a mean age of 55 years (SD 16) and a median of 58 (IQR 21 - 80) years. Anthropometric measurements showed a mean weight of 81.7 kg (SD 21.4) and a mean height of 1.71 m (SD 0.11), resulting in a mean BMI of 27.7 (SD 6.0) and a median BMI of 27.0 (IQR 17.3 - 42.9). According to World Health Organization BMI classifications, two patients (4%) were underweight, 17 (34%) had normal weight, 19 (38%) were overweight, and 12 (24%) were classified as obese; 5 had type I obesity (BMI 30 - 34.9), 5 had type II obesity (BMI 35 - 39.9), and 2 had type III obesity (BMI ≥40) (Table S1 in Multimedia Appendix 5).

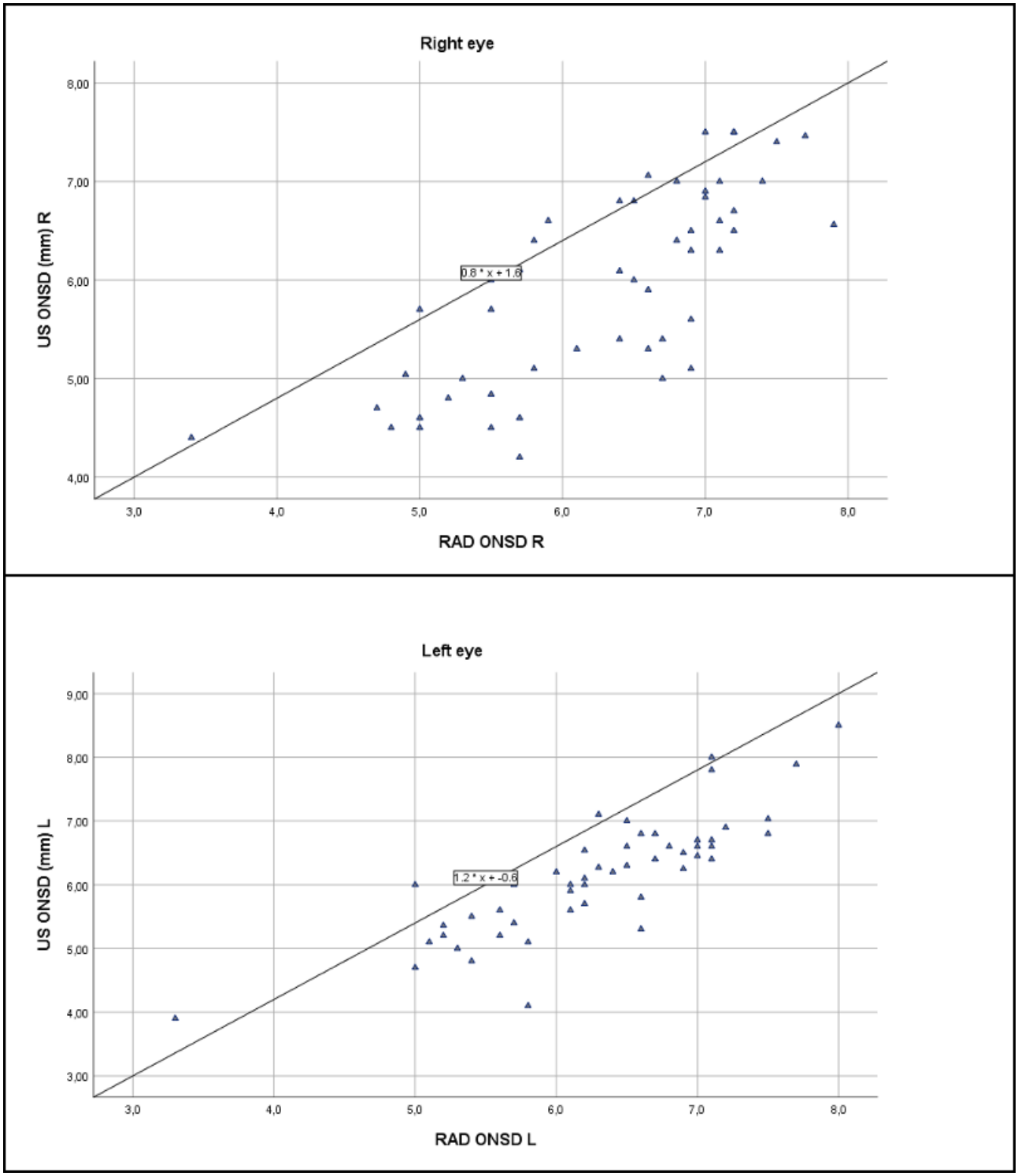
### ONSD Measurements

The average ONSD measured by ultrasound (US) was 5.94 mm (SD 0.99) in the right eye and 6.15 mm (SD 0.94) in the left eye. In comparison, radiological assessment (RAD) showed slightly higher values of 6.28 mm (SD 0.94) and 6.30 mm (SD 0.86) for the right and left eyes, respectively. Differences between the two methods ranged from –1 mm to 1.8 mm in the right eye and –1 mm to 1.7 mm in the left eye. Statistical analysis suggested a strong positive correlation between US and RAD measurements for both eyes, with correlation coefficients of 0.78 (right eye, Spearman) and 0.84 (left eye, Pearson), both highly significant (*P*<.001). These findings show a good agreement between ultrasound and radiological methods for assessing ONSD in patients undergoing brain tumor resection (Table S2 in Multimedia Appendix 6).

### Correlation Between Imaging Modalities

There was a clear, statistically significant correlation between ultrasound-measured ONSD values and radiological imaging values. In the right eye, the correlation coefficient was 0.779 (Spearman), and in the left eye, it was 0.836 (Pearson), with both showing *P* values below .001. These findings suggest that the two methods consistently align in their assessment of ONSD, with strong agreement in both eyes (Figure 1).

**Figure 1.** Optic nerve sheath diameter (ONSD) correlation with ultrasound (US) and magnetic resonance imaging (RAD) in patients scheduled for brain tumor resection (University Hospital, July 2021-May 2023, right and left eyes).




### Intraclass Correlation Coefficient

We used ICCs to assess the agreement between ultrasound and radiology measurements of ONSD. For the right eye, the ICC was 0.707 for individual measurements and 0.829 for averaged measurements. The agreement was even stronger in the left eye, with ICCs of 0.823 and 0.903 for single- and average-measures, respectively. All results were statistically significant (*P* values=.001), indicating strong agreement between the two methods. Full details of the ICC values and confidence intervals are available in Table S3 in Multimedia Appendix 7.

## Discussion

This study evaluated the relationship between ONSD measurements obtained via ultrasound and MRI in adult patients undergoing brain tumor resection. We found a strong and statistically significant correlation between the two methods, with minor average differences and high ICCs. These findings

suggest that ultrasound may be a practical alternative to MRI for assessing ONSD in the perioperative setting, where rapid, noninvasive evaluation of ICP is often required [20].

In addition to the primary findings, we observed that ultrasound-derived ONSD measurements were generally consistent with MRI-derived measurements, with most values falling within clinically acceptable limits of agreement. Notably, the measurements showed high reproducibility in both eyes, reinforcing the potential utility of ultrasound in routine practice when MRI is unavailable or impractical in patients with supratentorial masses.

Previous research has demonstrated strong associations between increased ONSD and raised ICP [21,22]. Our findings align with those reported by Bäuerle et al [23], who found a correlation between ultrasound- and MRI-based ONSD in a small cohort of healthy volunteers ($r$=0.72). However, their population was younger and nonsurgical. Kerscher et al [24-26] reported even stronger agreement in pediatric patients ($r$=0.976), but their findings are limited to that population. In contrast, our study focuses on adults in a surgical context and provides new evidence supporting the feasibility of ultrasound-assessed ONSD in preoperative neurosurgical care.

A key strength of our approach was the use of POCUS performed by anesthesiologists familiar with the technique. While the clinicians involved had prior ultrasound experience, ONSD-specific training was brief, which reflects the method's accessibility. Notably, even minimal training has been shown to produce reliable measurements, as seen in studies involving nonphysician trainees [27]. This highlights the practicality of implementing ONSD ultrasound in various clinical settings, especially where MRI is unavailable or delayed.

It is also worth noting that while ONSD ultrasound is most often discussed in the context of trauma or critical care [28], our findings suggest its potential value in elective neurosurgery. In patients with supratentorial tumors, where intraoperative ICP changes can have serious consequences, the ability to assess ONSD in real time could support timely decision-making. Given the growing interest in using POCUS for perioperative neuromonitoring, these findings may help guide future protocols, especially in resource-limited environments.

However, several limitations must be acknowledged. First, the study did not assess interobserver variation, which limits conclusions about the method's broader applicability to other clinicians or settings. Second, although trained anesthesiologists performed the ultrasound scans, access to formal ultrasound education may not be universal, especially in remote or low-resource centers. Third, our sample included only patients with tumoral disease; thus, the findings may not extend to other intracranial pathologies, such as hemorrhage or infection. Finally, the study's sample size, while larger than some previous comparisons, remains relatively small and limits the strength of conclusions that can be drawn.

Despite these limitations, this study provides valuable insight into the use of ultrasound for optic nerve evaluation in neurosurgical patients. By demonstrating a strong correlation with MRI, our findings support further exploration of ONSD ultrasound as a screening or monitoring tool for raised ICP, particularly in settings where MRI is unavailable or impractical. Recent open access articles have also pointed out the need for more consistent training and standardized ONSD protocols [32, which aligns with our recommendation for larger, coordinated studies to refine measurement techniques.

Future studies should focus on standardizing training protocols, assessing consistency between different examiners, and evaluating ONSD thresholds across a broader range of intracranial conditions. Larger, multicenter trials could help define normal and pathological ONSD values more precisely and assess how ultrasound-guided ICP screening might impact clinical outcomes. As interest in perioperative neuromonitoring continues to grow, incorporating tools like ONSD ultrasound may help streamline care, improve responsiveness to neurological changes, and broaden access to noninvasive ICP monitoring.

## Funding

The authors declared no financial support was received for this work.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Optic nerve sheath.
[DOCX File, 289 KB - periop_v9i1e67480_app1.docx ]

Multimedia Appendix 2
Optic nerve measurement 5 mm below retinal layer as indicated in the picture. This is the level to measure the transverse diameter of the optic nerve.
[DOCX File, 292 KB - periop_v9i1e67480_app2.docx ]

Multimedia Appendix 3
Optic nerve diameter measurement.

[DOCX File, 299 KB - periop_v9i1e67480_app3.docx ]

Multimedia Appendix 4
Optic nerve magnetic resonance imaging measurement.
[DOCX File, 483 KB - periop_v9i1e67480_app4.docx ]

Multimedia Appendix 5
Optic nerve diameter correlation with ultrasound and magnetic resonance imaging. University Hospital, July 2021-May 2023.
[PNG File, 88 KB - periop_v9i1e67480_app5.png ]

Multimedia Appendix 6
Optic nerve diameter correlation with ultrasound (US) and magnetic resonance imaging (RAD) in patients scheduled for brain tumor resection. University Hospital, July 2021-May 2023.
[DOCX File, 15 KB - periop_v9i1e67480_app6.docx ]

Multimedia Appendix 7
Interclass correlation coefficient.
[PNG File, 77 KB - periop_v9i1e67480_app7.png ]

## References


1. Byrne SF, Green RL. Ultrasound of the Eye and Orbit, 2nd edition: Mosby; 2002.
2. Bedi DG, Gombos DS, Ng CS, Singh S. Sonography of the eye. AJR Am J Roentgenol 2006 Oct;187(4):1061-1072. [doi: 10.2214/AJR.04.1842] [Medline: 16985158]
3. Shirodkar CG, Munta K, Rao SM, Mahesh MU. Correlation of measurement of optic nerve sheath diameter using ultrasound with magnetic resonance imaging. Indian J Crit Care Med 2015 Aug;19(8):466-470. [doi: 10.4103/0972-5229.162465] [Medline: 26321806]
4. Ohle R, McIsaac SM, Woo MY, Perry JJ. Sonography of the optic nerve sheath diameter for detection of raised intracranial pressure compared to computed tomography: a systematic review and meta-analysis. J Ultrasound Med 2015 Jul;34(7):1285-1294. [doi: 10.7863/ultra.34.7.1285] [Medline: 26112632]
5. Liu D, Kahn M. Measurement and relationship of subarachnoid pressure of the optic nerve to intracranial pressures in fresh cadavers. Am J Ophthalmol 1993 Nov 15;116(5):548-556. [doi: 10.1016/s0002-9394(14)73195-2] [Medline: 8238213]
6. Hayreh SS. Pathogenesis of optic disc edema in raised intracranial pressure. Prog Retin Eye Res 2016 Jan;50:108-144. [doi: 10.1016/j.preteyeres.2015.10.001] [Medline: 26453995]
7. Newman WD, Hollman AS, Dutton GN, Carachi R. Measurement of optic nerve sheath diameter by ultrasound: a means of detecting acute raised intracranial pressure in hydrocephalus. Br J Ophthalmol 2002 Oct;86(10):1109-1113. [doi: 10.1136/bjo.86.10.1109] [Medline: 12234888]
8. Moretti R, Pizzi B, Cassini F, Vivaldi N. Reliability of optic nerve ultrasound for the evaluation of patients with spontaneous intracranial hemorrhage. Neurocrit Care 2009 Dec;11(3):406-410. [doi: 10.1007/s12028-009-9250-8] [Medline: 19636971]
9. Robba C, Santori G, Czosnyka M, et al. Optic nerve sheath diameter measured sonographically as non-invasive estimator of intracranial pressure: a systematic review and meta-analysis. Intensive Care Med 2018 Aug;44(8):1284-1294. [doi: 10.1007/s00134-018-5305-7] [Medline: 30019201]
10. Rajajee V, Vanaman M, Fletcher JJ, Jacobs TL. Optic nerve ultrasound for the detection of raised intracranial pressure. Neurocrit Care 2011 Dec;15(3):506-515. [doi: 10.1007/s12028-011-9606-8] [Medline: 21769456]
11. Robba C, Donnelly J, Cardim D, et al. Optic nerve sheath diameter ultrasonography at admission as a predictor of intracranial hypertension in traumatic brain injured patients: a prospective observational study. J Neurosurg 2020 Apr 1;132(4):1279-1285. [doi: 10.3171/2018.11.JNS182077] [Medline: 30849751]
12. Bekerman I, Gottlieb P, Vaiman M. Variations in eyeball diameters of the healthy adults. J Ophthalmol 2014;2014:503645. [doi: 10.1155/2014/503645] [Medline: 25431659]
13. Chandrapatham K, Cardim D, Czosnyka M, et al. Variability of the optic nerve sheath diameter on the basis of sex and age in a cohort of healthy volunteers. Acta Neurochir Suppl 2021;131:121-124. [doi: 10.1007/978-3-030-59436-7_25] [Medline: 33839831]
14. Abbinante G, Vitiello L, Coppola A, Salerno G, Gagliardi V, Pellegrino A. Optic nerve ultrasound evaluation in children: a review. Diagnostics (Basel) 2023 Feb 1;13(3):535. [doi: 10.3390/diagnostics13030535] [Medline: 36766639]
15. Betcher J, Becker TK, Stoyanoff P, Cranford J, Theyyunni N. Military trainees can accurately measure optic nerve sheath diameter after a brief training session. Mil Med Res 2018 Dec 20;5(1):42. [doi: 10.1186/s40779-018-0189-y] [Medline: 30572931]
16. Kimberly HH, Shah S, Marill K, Noble V. Correlation of optic nerve sheath diameter with direct measurement of intracranial pressure. Acad Emerg Med 2008 Feb;15(2):201-204. [doi: 10.1111/j.1553-2712.2007.00031.x] [Medline: 18275454]

17. Shafiq F, Ali MA, Khan FA. Pre-anaesthetic assessment of intracranial pressure using optic nerve sheath diameter in patients scheduled for elective tumour craniotomy. J Ayub Med Coll Abbottabad 2018;30(2):151-154. [Medline: 29938408]
18. Stevens RRF, Gommer ED, Aries MJH, et al. Optic nerve sheath diameter assessment by neurosonology: a review of methodologic discrepancies. J Neuroimaging 2021 Sep;31(5):814-825. [doi: 10.1111/jon.12906] [Medline: 34270144]
19. Trocha G, Bonilla A, Romero C, Palacios J, Molano-Gonzalez N. Ultrasound measurement of optic nerve sheath diameter in a healthy adult Colombian population. BMC Neurol 2023 Jan 13;23(1):16. [doi: 10.1186/s12883-023-03062-4] [Medline: 36639617]
20. Jiang J, Xie X. Ultrasound measurement of optic nerve sheath diameter for intracranial pressure monitoring. Interdiscip Neurosurg 2025 Sep;41:102091. [doi: 10.1016/j.inat.2025.102091]
21. Desai VR, Sadrameli SS, Hoppe S, et al. Contemporary management of increased intraoperative intracranial pressure: evidence-based anesthetic and surgical review. World Neurosurg 2019 Sep;129:120-129. [doi: 10.1016/j.wneu.2019.05.224] [Medline: 31158533]
22. Padayachy LC, Padayachy V, Galal U, Pollock T, Fieggen AG. The relationship between transorbital ultrasound measurement of the optic nerve sheath diameter (ONSD) and invasively measured ICP in children.: Part II: age-related ONSD cut-off values and patency of the anterior fontanelle. Childs Nerv Syst 2016 Oct;32(10):1779-1785. [doi: 10.1007/s00381-016-3068-4] [Medline: 27659820]
23. Bäuerle J, Schuchardt F, Schroeder L, Egger K, Weigel M, Harloff A. Reproducibility and accuracy of optic nerve sheath diameter assessment using ultrasound compared to magnetic resonance imaging. BMC Neurol 2013 Dec 1;13:187. [doi: 10.1186/1471-2377-13-187] [Medline: 24289136]
24. Kerscher SR, Schöni D, Hurth H, et al. The relation of optic nerve sheath diameter (ONSD) and intracranial pressure (ICP) in pediatric neurosurgery practice - part I: correlations, age-dependency and cut-off values. Childs Nerv Syst 2020 Jan;36(1):99-106. [doi: 10.1007/s00381-019-04266-1] [Medline: 31256241]
25. Kerscher SR, Schöni D, Neunhoeffer F, et al. The relation of optic nerve sheath diameter (ONSD) and intracranial pressure (ICP) in pediatric neurosurgery practice - Part II: influence of wakefulness, method of ICP measurement, intra-individual ONSD-ICP correlation and changes after therapy. Childs Nerv Syst 2020 Jan;36(1):107-115. [doi: 10.1007/s00381-019-04336-4] [Medline: 31392457]
26. Kerscher SR, Zipfel J, Groeschel S, Bevot A, Haas-Lude K, Schuhmann MU. Comparison of B-scan ultrasound and mri-based optic nerve sheath diameter (ONSD) measurements in children. Pediatr Neurol 2021 Nov;124:15-20. [doi: 10.1016/j.pediatrneurol.2021.08.002] [Medline: 34508997]
27. Tayal VS, Neulander M, Norton HJ, Foster T, Saunders T, Blaivas M. Emergency department sonographic measurement of optic nerve sheath diameter to detect findings of increased intracranial pressure in adult head injury patients. Ann Emerg Med 2007 Apr;49(4):508-514. [doi: 10.1016/j.annemergmed.2006.06.040] [Medline: 16997419]
28. Herrero MP, Fernández FDP, Moreno MM, Martín AM, Fernández-Caballero S. Indirect measurement of perioperative intracranial pressure-eye and orbitary ultrasound utility. Advances in Medicine and Engineering Interdisciplinary Research 2024;2(3):78-85. [doi: 10.32629/ameir.v2i3.2615]


## Abbreviations

**CT:** computed tomography
**ETD:** transverse diameter of the eyeball
**ICC:** intraclass correlation coefficient
**ICP:** intracranial pressure
**MRI:** magnetic resonance imaging
**ONSD:** optic nerve sheath diameter
**POCUS:** point-of-care ultrasound
**RAD:** radiological assessment
**US:** ultrasonography


*Edited by N Rohatgi; submitted 17.Nov.2024; peer-reviewed by D Janitschke, MI Hirzallah, T Aslanidis; revised version received 16.Jan.2026; accepted 19.Jan.2026; published 17.Apr.2026.*

*Please cite as:*
*Giraldo M, Lopera LM, Balbaa A, Gonzalez N, Cezar Zamper RP, Mayich M, Boulton M*
*Comparison Between Ultrasound and Magnetic Resonance Imaging Measurements of the Optic Nerve Sheath Diameter in Patients Undergoing Intracranial Surgery: Prospective Observational Single-Center Study*
*JMIR Perioper Med 2026;9:e67480*
*URL: https://periop.jmir.org/2026/1/e67480*
*doi:10.2196/67480*

# Survival Prediction in Patients With Bladder Cancer Undergoing Radical Cystectomy Using a Machine Learning Algorithm: Retrospective Single-Center Study

Francesco Andrea Causio[1,2], MD; Vittorio De Vita[1,2], MD; Andrea Nappi[1,3], BSc; Melissa Sawaya[4], PhD; Bernardo Rocco[5,6], MD, PhD; Nazario Foschi[5,6], MD; Giuseppe Maioriello[5,6], MD; Pierluigi Russo[1,5,6,7], MD

[1]Italian Society for Artificial Intelligence in Medicine (SIIAM - Società Italiana Intelligenza Artificiale in Medicina), Rome, Italy

[2]University Department of Life Sciences and Public Health, Section of Hygiene, Università Cattolica del Sacro Cuore, Largo F Vito 1, Rome, Italy

[3]Computer Science Department, University of Twente, Twente, The Netherlands

[4]Université Paris-Saclay, UVSQ, Inserm, Gustave Roussy, CESP, Villejuif, France

[5]University Department of Urology, Università Cattolica del Sacro Cuore, Rome, Italy

[6]University Department Medicine and Translational Surgery, Università Cattolica del Sacro Cuore, Rome, Italy

[7]Department of Life Science, Health, and Health Professions, Università Degli Studi Link, Rome, Italy

**Corresponding Author:**
Vittorio De Vita, MD
Italian Society for Artificial Intelligence in Medicine (SIIAM - Società Italiana Intelligenza Artificiale in Medicina), Rome, Italy

## Abstract

**Background:** Traditional statistical models often fail to capture the complex dynamics influencing survival outcomes in patients with bladder cancer after radical cystectomy, a procedure where approximately 50% of patients develop metastases within 2 years. The integration of artificial intelligence (AI) offers a promising avenue for enhancing prognostic accuracy and personalizing treatment strategies.

**Objective:** This study aimed to develop and evaluate a machine learning algorithm for predicting disease-free survival (DFS), overall survival (OS), and the cause of death in patients with bladder cancer undergoing cystectomy, using a comprehensive dataset of clinical and pathological variables.

**Methods:** Retrospective data of 370 patients with bladder cancer who underwent radical cystectomy at Fondazione Policlinico Universitario Agostino Gemelli IRCCS, Rome, Italy, were collected. The dataset comprised 20 input variables, encompassing demographics, tumor characteristics, treatment variables, and inflammatory markers. For specific analyses and models, we used patient subcohorts. The CatBoost algorithm was used for regression tasks (DFS in 346 patients, OS in 347 patients) and a binary classification task (tumor-related death in 312 patients). Model performance was assessed using mean absolute error (MAE) for regression and $F_1$-score for classification, prioritizing a minimum recall of 75% for tumor-related deaths. Five-fold cross-validation and Shapley additive explanations (SHAP) values were used to ensure robustness and interpretability.

**Results:** For DFS prediction, the CatBoost model achieved an MAE of 18.68 months, with clinical tumor stage and pathological tumor classification identified as the most influential predictors. OS prediction yielded an MAE of 17.2 months, which improved to 14.6 months after feature filtering, where tumor classification and the systemic immune-inflammation index (SII) were most impactful. For tumor-related death classification, the model achieved a recall of 78.6% and an $F_1$-score of 0.44 for the positive class (tumor-related deaths), correctly identifying 11 of 14 cases. Bladder tumor position was the most influential feature for cause-of-death prediction.

**Conclusions:** The developed machine learning algorithm demonstrates promising accuracy in predicting survival and the cause of death in patients with bladder cancer after cystectomy. The key predictors include clinical and pathological tumor staging, systemic inflammation (SII), and bladder tumor position. These findings highlight the potential of AI in providing clinicians with an objective, data-driven tool to improve personalized prognostic assessment and guide clinical decision-making.

***(JMIR Perioper Med 2026;9:e86666)*** doi:10.2196/86666

**KEYWORDS**

cystectomy; disease-free survival; artificial intelligence; neoplasm staging; retrospective studies; urinary bladder neoplasms; clinical decision-making; machine learning; statistical models

## Introduction

In the evolving landscape of health care, the integration of artificial intelligence (AI) into clinical decision-making has gained significant momentum, particularly in the realm of oncology [1,2]. With advancements in machine learning techniques, health care professionals are increasingly harnessing the power of AI to enhance diagnosis, prognosis, and treatment planning. The exponential growth of digital health care data, including electronic health records, medical imaging, genomic data, and real-time patient monitoring, has fueled the development of predictive algorithms [1,3].

The field of urology is complex: cancerous conditions benefit from the leverage of additional data sources and decision-making algorithms that allow physicians to plan treatment while considering several complex factors. Urological cancers, including prostate, bladder, and renal cancers, place a considerable burden on health care systems worldwide [4]. These malignancies often require complex management involving early diagnosis, accurate staging, and personalized treatment strategies to optimize patient outcomes. Traditional methods of assessing prognosis rely heavily on statistical models that may not capture the multifaceted nature of cancer behavior and patient responses to treatment. Conventional regression statistics often fail to provide the depth of analysis required to address the complexities of cancer management. In contrast, AI techniques, such as artificial neural networks, Bayesian networks, and neuro-fuzzy modeling systems, offer innovative approaches to constructing data-driven models that can adapt to the heterogeneous nature of cancer [5].

The potential of AI in predicting patient outcomes is particularly evident in its ability to analyze large datasets without the constraints of predetermined statistical distributions. By leveraging retrospective data, we can develop algorithms that not only identify patterns and correlations but also provide insights into individual patient behavior. This capability is crucial for clinicians who face the challenge of tailoring treatment plans to the unique characteristics of each patient. In the context of mortality and postoperative survival, the application of AI can provide critical insights that enhance our understanding of patient outcomes following surgical interventions. The ability to predict which patients are at higher risk of complications or recurrence can lead to more informed clinical decisions, ultimately improving the quality of care [6]. For instance, machine learning algorithms can analyze a multitude of variables, including clinical, pathological, and demographic factors, to generate individualized risk profiles that guide treatment strategies and follow-up plans [7]. Recent urological research has shown that combining hematological inflammation indexes with machine learning algorithms can improve the prediction of surgical outcomes, as demonstrated in patients who underwent urethroplasty [8].

In this study, we focus specifically on the training of an AI algorithm using retrospective data collected from patients diagnosed with bladder cancer who underwent radical cystectomy. Patients with localized muscle-invasive or recurrent non–muscle-invasive bladder cancer benefit most from radical cystectomy, which may be preceded by neoadjuvant chemotherapy in selected cases, in terms of local disease control. Even with sufficient local control achieved through cystectomy, approximately 50% of patients develop metastases within 2 years and may ultimately die from the disease. This is likely due to the existence of regional or distant microscopic metastatic disease at the time of surgery [9]. The proposed methodology will involve the comprehensive examination of variables associated with patient demographics, tumor characteristics, treatment modalities, and postoperative outcomes. Using machine learning techniques, we aim to identify key predictors of mortality and postoperative survival, ultimately constructing a predictive model with potential relevance for clinical decision-making.

## Methods

### Study Design

We collected retrospective data on patients with high-risk and very high-risk non–muscle-invasive bladder cancer and muscle-invasive bladder cancer who underwent radical cystectomy at Fondazione Policlinico Universitario Agostino Gemelli IRCCS in Rome, Italy. The dataset included data on various clinical and pathological variables from 370 patients.

### Ethical Considerations

Ethical approval was obtained from the institutional ethical review board (protocol number 676 - 02). As primary consent for data collection covered secondary analyses, additional consent was not required for this study. The data used in this study were anonymized, and no compensation was provided to patients.

### Data Collection and Preprocessing

Clinical and pathological data were extracted from medical records, including demographic, lifestyle, tumor, treatment, and laboratory variables. The dataset was split into three outcome-specific subsets to maximize the usable sample for each task:

1. DFS dataset: predicting disease-free survival (DFS) in months (346 patients in total).
2. OS dataset: predicting overall survival (OS) in months (347 patients in total).
3. Death cause dataset: for classification purposes, the classes “death from other causes” and “alive” were merged into a single negative class to create a binary variable. Therefore, the cause is defined as either “no” or “yes,” depending on whether it was tumor related (312 patients in total).

The variables included in the dataset are detailed in Table 1. All categorical variables were cast as strings to allow native handling by CatBoost (version 1.2.8; Yandex).

A total of 20 input variables were selected for model development.

**Table .** Variables included in the study

| Variable (English) | Description | Data type | Input/output |
|---|---|---|---|
| Patient demographics and lifestyle | | | |
| AGE | Age (years) | Numerical | Input |
| BMI | Body mass index (kg/m$^2$) | Numerical | Input |
| SEX | Biological sex (0: man, 1: woman) | Categorical | Input |
| SMOKE | Smoker (0: no, 1: yes) | Categorical | Input |
| Patient medical history | | | |
| DM | Patient has diabetes mellitus (0: no, 1: yes) | Categorical | Input |
| PRIOR SURGERY | Patient had previously undergone surgery in the abdominal area (0: no, 1: yes) | Categorical | Input |
| PRIOR RADIOTHERAPY | Patient had previously received radiotherapy in the abdominal area (0: no, 1: yes) | Categorical | Input |
| PRIOR SYSTEMIC CHEMOTHERAPY | Patient had previously received systemic chemotherapy (0: no, 1: yes) | Categorical | Input |
| Tumor characteristic | | | |
| BLADDER TUMOR POSITION | Identifier of tumor position (0: intertrigonal zone, 1: right periosteal, 2: left periosteal, 3: dome, 4: posterior wall, 5: right lateral wall, 6: left lateral wall, 7: prostatic urethra, 8: anterior wall, 9: entire bladder, 10: bladder base) | Categorical | Input |
| TUMOR DIMENSION | Tumor dimension (cm) | Numerical | Input |
| PRE-HYDRONEPHROSIS | Hydronephrosis (0: no, 1: right hydronephrosis, 2: left hydronephrosis, 3: bilateral hydronephrosis) | Categorical | Input |
| H.E. TURV | Histological examination for transurethral resection of the bladder (0: localized to mucosa +/– submucosa multirecurrent, 1: muscle-invasive, 2: squamous) | Categorical | Input |
| LVI | Lymphovascular invasion (0: absent, 1: present) | Categorical | Input |
| CTS | Clinical tumor stage (0: cTa, 1: cTis, 2: cT1, 3: cT2, 4: cT3, 5: cT4) | Categorical | Input |
| TC | Tumor classification (1: T0, 2: Ta, 3: Tis, 4: T1, 5: T2a, 6: T2b, 7: T3a, 8: T3b, 9: T4a, 10: T4b) | Categorical | Input |
| Inflammatory and immune marker | | | |
| SII | Systemic immune-inflammation index (decimals) | Numerical | Input |
| Treatment and outcome | | | |
| UD | Urinary diversion type (0: Bricker ileal conduit, 1: ureterocutaneostomy, 2: vesicoileal pouch) | Categorical | Input |
| RECURRENCE | Tumor recurrence (0: no, 1: yes) | Categorical | Input |
| DFS | Disease-free survival after treatment (in months) | Numerical | Output |

| Variable (English) | Description | Data type | Input/output |
|---|---|---|---|
| OS | Overall survival: time from diagnosis/treatment start to death from any cause (in months) | Numerical | Output |
| DEATH CAUSE | Cause of death (X: alive, 1: other, 2: cancer); later merged (0: alive + other, 1: cancer) | Categorical | Output |

## Machine Learning Models

To predict clinical outcomes, we used the CatBoost algorithm for both regression (DFS and OS) and classification (cause of death) tasks, as it is effective for small and structured datasets.

For DFS and OS, we applied CatBoostRegressor models. For predicting the cause of death, we used the CatBoostClassifier, with the binary outcome of death being tumor related or not.

## DFS and OS Model Evaluation

For the regression tasks (DFS and OS), we evaluated model performance using mean absolute error (MAE) to quantify the average prediction error in months.

## Cause-of-Death Classification Model

For the classification task (cause of death), we evaluated performance using the $F_1$-score. The $F_1$-score is a single metric that balances precision and recall, particularly useful in cases of imbalanced classes where the positive class is of primary interest. For class 1 (tumor-related deaths), it was calculated as the harmonic mean of precision and recall:

(1)F1=2∗(Precision∗Recall)/(Precision+Recall)

Precision is the proportion of correctly identified positive predictions among all positive predictions, and recall is the proportion of correctly identified positive predictions among all actual positives.

(2)Precision=TP/(TP+FP)

(3)Recall=TP/(TP+FN)

Confusion matrices were used to examine prediction distributions, and probability thresholds were adjusted to optimize recall while limiting false positives. To account for class imbalance in the classification task, we applied custom class weights. We adjusted the decision threshold, aiming for a minimum recall of 75% to ensure that most tumor-related deaths were accurately identified and classified.

## Cross-Validation and Hyperparameter Tuning

All models were trained and evaluated using 5-fold cross-validation to ensure generalizability and reduce the risk of overfitting, especially given the relatively small dataset. In addition, we applied early stopping with a patience range of 30 to 50 rounds, allowing the model to terminate training once performance ceased to improve on the validation fold.

To enhance the interpretability and transparency of the developed machine learning models, we used violin plots and Shapley additive explanations (SHAP) scatterplots to investigate the impact of variables on the prediction of the results. Violin plots show the effect of each variable on the results, both in terms of direction (favorable or unfavorable) and intensity. The SHAP scatterplot assigns an importance value to each feature for a particular prediction. For each patient, the SHAP values revealed the specific features driving the predicted risk of tumor death. By aggregating the SHAP values across the entire cohort, the overall impact and importance of each clinical and pathological variable on the model's outcome predictions were determined. This enabled the identification of the most significant factors influencing the outcomes of patients with bladder cancer after cystectomy.

This paper presents only the most significant results of the analysis. The complete analysis is available online for open access [10].

## Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis Guidelines for AI

To enhance the transparency, interpretability, and reproducibility of our machine learning-based prediction models, this study adheres to the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) statement, specifically considering the extensions for AI (TRIPOD+AI). The TRIPOD+AI guidelines provide a standardized framework for reporting studies that develop or validate prediction models, ensuring that sufficient detail is provided for critical appraisal and replication by other researchers. By following these guidelines, we aim to clearly articulate the study design, data characteristics, model development process, and performance evaluation, thereby contributing to the responsible and rigorous application of AI in medical research (Table 2).

**Table .** Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis checklist for reporting studies involving artificial intelligence (TRIPOD+AI).

| TRIPOD+AI item | Description of reporting in this study |
|---|---|
| 1. Title | • Survival Prediction in Patients with Bladder Cancer Undergoing Radical Cystectomy Using a Machine Learning Algorithm: Retrospective Single-Center Study |
| 2. Abstract | • The abstract summarizes the study's objectives, methods, key findings, and conclusions. |
| 3. Introduction - background | • The introduction will establish the clinical context of bladder cancer, the prognostic challenges, and the rationale for using machine learning. |
| 4. Methods - participants | |
| 4a. Eligibility criteria | • Patients who underwent radical cystectomy for bladder cancer, with available data for selected variables |
| 4b. Settings and locations | • Data collected retrospectively from a single institution: Fondazione Policlinico Universitario Agostino Gemelli IRCCS, Rome, Italy |
| 4c. Source of data | • Patient medical records |
| 5. Data acquisition method | • Retrospective data extraction into a spreadsheet |
| 6. Methods - outcome | |
| 6a. Definition of outcomes | • Disease-free survival (DFS): time in months from treatment to recurrence or death (event) or last follow-up (censored)<br>• Overall survival (OS): time in months from diagnosis/treatment start to death from any cause (event) or last follow-up (censored)<br>• DEATH CAUSE: binary classification (0: did not die from the tumor, 1: died from the tumor) |
| 6b. Outcome measurement | • DFS and OS were calculated from documented dates.<br>• The cause of death was extracted from medical records and recategorized for binary classification. |
| 7. Methods - predictors | |

| TRIPOD+AI item | Description of reporting in this study |
|---|---|
| 7a. Definition of all predictors | • AGE: patient's age (years)<br>• BMI (kg/m$^2$)<br>• DM: patient has diabetes mellitus (0: no, 1: yes)<br>• PRIOR SURGERY: patient had previously undergone surgery in the abdominal area (0: no, 1: yes)<br>• PRIOR RADIOTHERAPY: patient had previously received radiotherapy in the abdominal area (0: no, 1: yes)<br>• PRIOR SYSTEMIC CHEMOTHERAPY: patient had previously received systemic chemotherapy (0: no, 1: yes)<br>• BLADDER TUMOR POSITION: identifier of tumor position (0: intertrigonal zone, 1: right periosteal, 2: left periosteal, 3: dome, 4: posterior wall, 5: right lateral wall, 6: left lateral wall, 7: prostatic urethra, 8: anterior wall, 9: entire bladder, 10: bladder base)<br>• TUMOR DIMENSION: tumor dimension (cm)<br>• PRE-HYDRONEPHROSIS: pretreatment hydronephrosis (0: no, 1: right hydronephrosis, 2: left hydronephrosis, 3: bilateral hydronephrosis)<br>• SEX: biological sex (0: man, 1: woman)<br>• SMOKE: patient smokes (0: no, 1: yes)<br>• H.E. TURV: histological examination for transurethral resection of the bladder (0: localized to mucosa +/– submucosa multirecurrent, 1: muscle-invasive, 2: squamous)<br>• SII: systemic immune-inflammation index (decimals)<br>• UD: urinary diversion type (0: Bricker ileal conduit, 1: ureterocutaneostomy, 2: vesicoileal pouch)<br>• LVI: lymphovascular invasion (0: absent, 1: present)<br>• CTS: clinical tumor stage (0: cTa, 1: cTis, 2: cT1, 3: cT2, 4: cT3, 5: cT4)<br>• TC: tumor classification (1: T0, 2: Ta, 3: Tis, 4: T1, 5: T2a, 6: T2b, 7: T3a, 8: T3b, 9: T4a, 10: T4b) |
| 7b. Predictor measurement | • Predictors were measured clinically (eg, age and BMI), derived from patient history (eg, prior surgeries and smoking status), or derived from laboratory or pathology reports (eg, SII, tumor dimension, LVI, CTS, and TC). |
| 8. Methods - sample size | |
| 8a. Sample size determination | • The available retrospective data determined the sample size. No formal power calculation was performed due to the exploratory nature of the study and the limitations of the data. |
| 9. Methods - data handling | |
| 9a. Handling of missing data | • Rows with null values in specific critical variables ("TUMOR DIMENSION," "LVI," "TC," "H.E. TURV," "RECURRENCE," "DFS," "OS," "DEATH CAUSE") were removed. No imputation was performed. |
| 9b. Data transformation | • Numerical variables were type-adjusted to int or float. Categorical variables were explicitly converted to category type. "DEATH CAUSE" was recategorized into a binary format. |
| 10. Methods - model development | |

| TRIPOD+AI item | Description of reporting in this study |
|---|---|
| 10a. Model type | • CatBoostRegressor (for DFS and OS) and CatBoostClassifier (for DEATH CAUSE). |
| 10b. Candidate predictors | • All 17 selected independent variables were used as candidate predictors for each model, based on the relevant dataset (df1 [DFS], df2 [OST], df3 [DEATH CAUSE]). |
| 10c. Handling of continuous predictors | • Continuous predictors (AGE, BMI, TUMOR DIMENSION, SII) were used directly by CatBoost, which handles them internally. |
| 10d. Handling of categorical predictors | • Categorical predictors were identified and explicitly converted to string type before training. CatBoost natively handles categorical features without explicit one-hot encoding. |
| 10e. Details of model fitting | • CatBoostRegressor (iterations=1000, learning_rate=0.05, depth=6, loss_function=”RMSE,” eval_metric=”MAE,” early_stopping_rounds=50, random_seed=42, verbose=0). Similar configurations for CatBoostClassifier, with “Logloss” or “MultiClass” as loss function. |
| 10f. Internal validation method | • Data were split into training (80%) and testing (20%) sets using train_test_split with random_state=42. Five-fold cross-validation (KFold, shuffle=True, random_state=42) was performed on the training set. |
| 10g. Performance metrics | • Regression (DFS, OS): mean absolute error (MAE)<br>• Classification (DEATH CAUSE): $F_1$-score for class 1 (tumor-related deaths), prioritizing recall |
| 11. Assessment of prediction performance | • Performance was assessed on the independent test set. For classification, a confusion matrix was used. |
| 12. Model interpretation methods | • CatBoost's built-in feature importance was used. Shapley additive explanations (SHAP) values were computed and visualized using violin plots and scatterplots to understand the individual contributions of each feature. |

# Results

## Patient Characteristics

The study included a final cohort of 374 patients. After excluding incomplete records, the analytical sample sizes were 346 for DFS prediction, 347 for OS prediction, and 312 for death cause prediction. Some records indicate fewer than 374 patients, as not all characteristics were available for every individual in the population.

Table 3 presents the baseline clinical, pathological, and demographic characteristics of the study population, comprising 79.4% (297/374) men and 20.6% (77/374) women. A majority, 79.4% (296/373) were active smokers, and 21.4% (80/374) had a diagnosis of diabetes mellitus. Prior surgery was reported for 33.7% (126/374) of patients, while previous radiotherapy and systemic chemotherapy were less common, 5.1% (19/374) and 3.5% (13/374), respectively. Preoperative hydronephrosis was present in approximately one-third of cases, most frequently unilaterally.

**Table .** Characteristics of the patients in the dataset.

| Characteristic | Value |
|---|---|
| Continuous variables | |
| Age (years), mean (SD) | 75.2 (9.5) |
| BMI (kg/m$^2$), mean (SD) | 26.6 (4.2) |
| Tumor dimension (cm), median (IQR) | 2.2 (1.1 - 2.8) |
| SII[a], median (IQR) | 654.7 (408.0 - 1047.0) |
| DFS[b] (months), median (IQR) | 23.0 (6.0 - 52.8) |
| OS[c] (months), median (IQR) | 29.0 (10.8 - 55.4) |
| Categorical variables, n/N (%) | |
| Sex | |
| Men | 297/374 (79.4) |
| Women | 77/374 (20.6) |
| Smoking status | |
| No | 77/373 (20.6) |
| Yes | 296/373 (79.4) |
| Diabetes mellitus | |
| No | 294/374 (78.6) |
| Yes | 80/374 (21.4) |
| Prior surgery | |
| No | 248/374 (66.3) |
| Yes | 126/374 (33.7) |
| Prior radiotherapy | |
| No | 355/374 (94.9) |
| Yes | 197/374 (5.1) |
| Prior chemotherapy | |
| No | 361/374 (96.5) |
| Yes | 13/374 (3.5) |
| Pre-hydronephrosis[d] | |
| None | 257/374 (68.7) |
| Right | 44/374 (11.8) |
| Left | 40/374 (10.7) |
| Bilateral | 33/374 (8.8) |
| Histological examination (H.E. TURV) | |
| Localized | 138/372 (37.1) |
| Muscle-invasive | 212/372 (57) |
| Squamous | 22/372 (5.9) |
| Urinary diversion type | |
| Bricker ileal conduit | 278/373 (74.5) |
| Ureterocutaneostomy | 36/373 (9.7) |
| Vesicoileal pouch | 59/373 (15.8) |
| Lymphovascular invasion | |
| Absent | 158/371 (42.6) |

| Characteristic | Value |
|---|---|
| Present | 213/371 (57.4) |
| Clinical tumor stage | |
| cTa | 132/366 (36.1) |
| cTis | 43/366 (11.7) |
| cT1 | 77/366 (21) |
| cT2 | 74/366 (20.2) |
| cT3 | 31/366 (8.5) |
| cT4 | 9/366 (2.5) |
| Tumor classification | |
| T0 | 16/342 (4.7) |
| Ta | 18/342 (5.3) |
| Tis | 60/342 (17.5) |
| T1 | 47/342 (13.7) |
| T2a | 54/342 (15.8) |
| T2b | 7/342 (2) |
| T3a | 80/342 (23.4) |
| T3b | 10/342 (2.9) |
| T4a | 42/342 (12.3) |
| T4b | 8/342 (2.4) |
| Bladder tumor position | |
| Intertrigonal zone | 44/365 (12.1) |
| Right periosteal | 25/365 (6.8) |
| Left periosteal | 39/365 (10.7) |
| Dome | 22/365 (6) |
| Posterior wall | 58/365 (15.9) |
| Right lateral wall | 71/365 (19.5) |
| Left lateral wall | 58/365 (15.9) |
| Prostatic urethra | 8/365 (2.2) |
| Anterior wall | 26/365 (7.1) |
| Entire bladder | 12/365 (3.3) |
| Bladder base | 2/365 (0.5) |
| Cause of death | |
| Alive | 205/363 (56.5) |
| Other | 71/363 (19.6) |
| Cancer | 87/363 (23.9) |

[a]SII: systemic immune-inflammation index.

[b]DFS: disease-free survival.

[c]OS: overall survival.

[d]Pre-hydronephrosis: pretreatment hydronephrosis.

Histologically, 57% (212/372) of tumors were muscle invasive, while 37.1% (138/372) were localized to the mucosa or submucosa, and 5.9% (22/372) exhibited squamous features. The most common urinary diversion method was Bricker ileal conduit (278/373, 74.5%), followed by vesicoileal pouch construction (59/373, 15.8%) and ureterocutaneostomy (36/373, 9.7%). Lymphovascular invasion was observed in 57.4% (213/371) of patients.

In terms of staging, the most frequent clinical tumor stages were cTa (132/366, 36.1%) and cT1 (77/366, 21%), while advanced

stages (cT3 and cT4) were less common (40/366, 11%). Tumor classification was heterogeneous, with T3a (80/342, 23.4%) and Tis (60/342, 17.5%) being most prevalent.

Regarding tumor location, the most frequent sites were the right lateral wall (71/365, 19.5%), the posterior wall (58/365, 15.9%), and the left lateral wall (58/365, 15.9%). At the time of data collection, of 363 patients, 205 (56.5%) were alive, 87 (24%) had died due to cancer-related causes, and 71 (19.6%) had died from other causes.

## DFS Prediction

The CatBoostRegressor model was trained to predict DFS in months. Input variables are indicated in Table 1. After 5-fold cross-validation and manual hyperparameter tuning, the model achieved an MAE of 18.68 months, indicating that, on average, the model's predictions deviated by approximately 1.5 years from the observed DFS.

Figure 1 presents the global feature importance ranking from the CatBoost model trained to predict DFS. This ranking reflects the contribution of each variable to reducing the model's prediction across all patients. The most influential predictor was the clinical tumor stage, with an importance score of approximately 17, followed by the pathological tumor classification, with an importance score of approximately 14, reflecting the role of tumor invasiveness, local extension, and accurate tumor staging in DFS prediction. The systemic immune-inflammation index (SII) ranked highly, with a score of approximately 9.5. To a lesser extent, demographic and anatomical variables, such as BMI, age, and tumor dimension, also contributed to the model.

**Figure 1.** Global feature importance ranking to predict disease-free survival. CTS: clinical tumor stage; DM: diabetes mellitus; H.E. TURV: histological examination for transurethral resection of the bladder; LVI: lymphovascular invasion; PRE-HYDRONEPHROSIS: pretreatment hydronephrosis; SII: systemic immune-inflammation index; TC: tumor classification; UD: urinary diversion type.

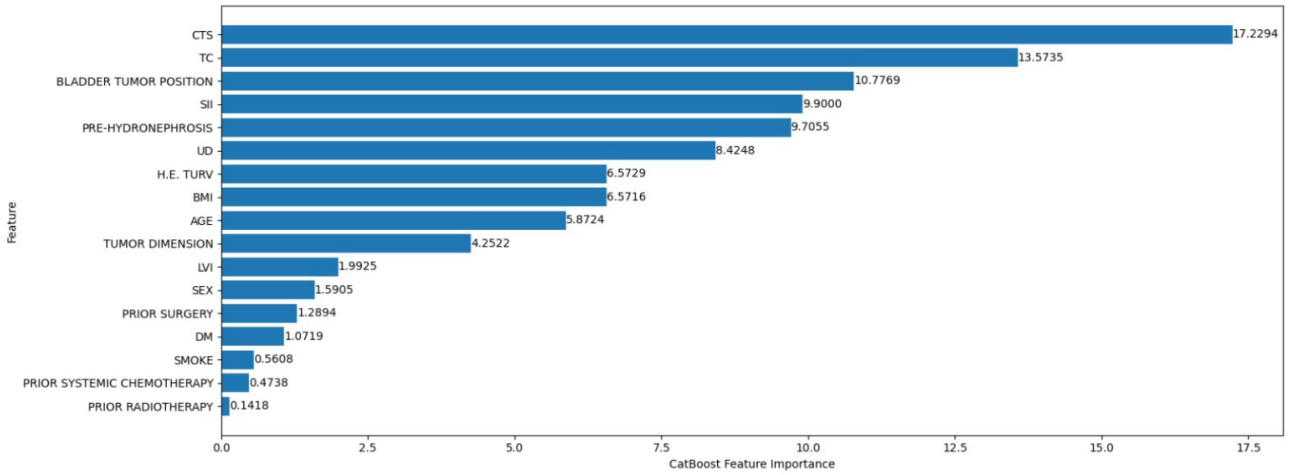



Figure 2 displays the SHAP summary plot for the DFS model, illustrating the distribution and direction of impact of each feature on the predicted DFS across all patients. Clinical tumor stage and tumor classification exhibited the widest distribution of SHAP values, confirming their dominant influence, where the predicted DFS substantially increased or decreased depending on their values. SII displayed a more balanced distribution, with both positive and negative effects depending on the value. In contrast, features such as prior treatment (eg, surgery, radiotherapy, and chemotherapy) and lifestyle factors (eg, smoking status and diabetes) had a SHAP distribution clustered near 0, indicating limited predictive power.

**Figure 2.** Violin plot of feature influence on disease-free survival prediction from the Shapley additive explanations (SHAP) analysis. CTS: clinical tumor stage; DM: diabetes mellitus; H.E. TURV: histological examination for transurethral resection of the bladder; LVI: lymphovascular invasion; PRE-HYDRONEPHROSIS: pretreatment hydronephrosis; SII: systemic immune-inflammation index; TC: tumor classification; UD: urinary diversion type.

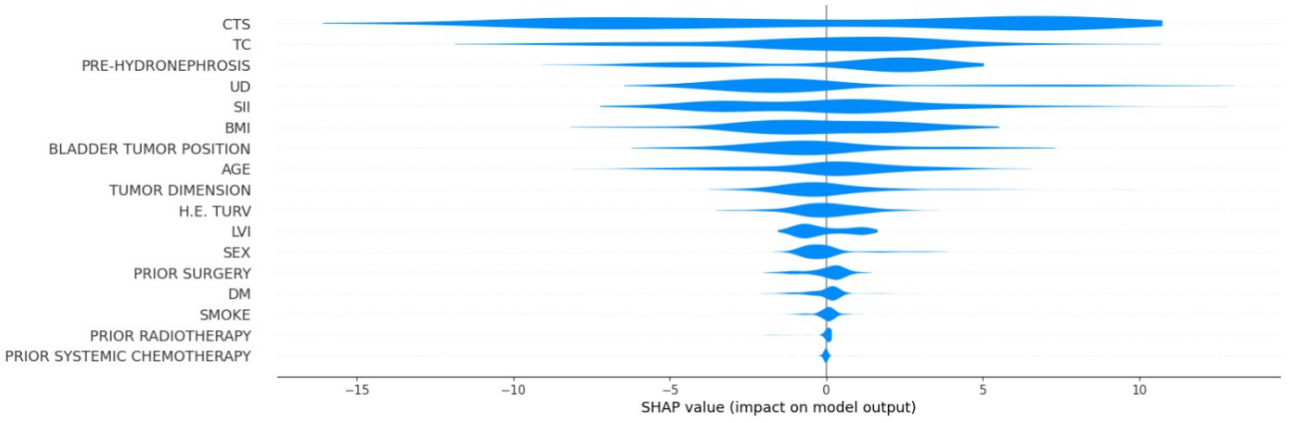
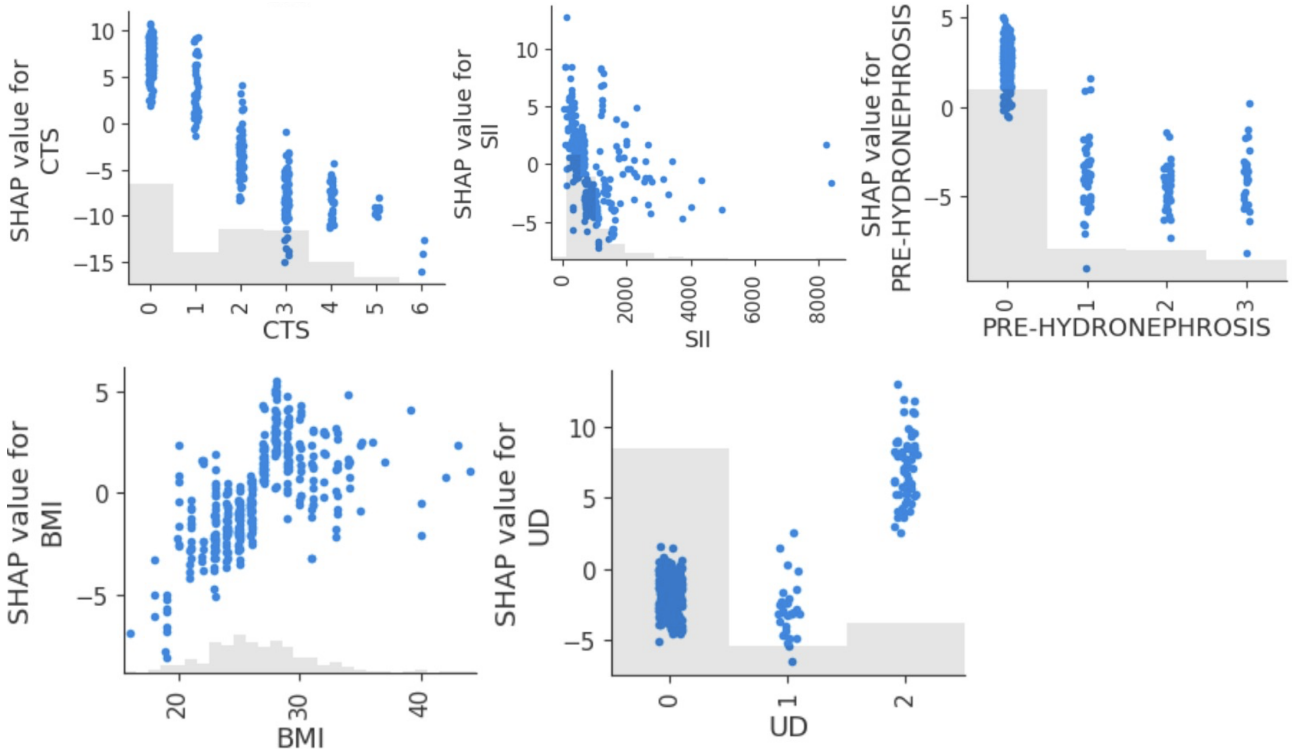


Figure 3 presents the SHAP dependence plots for 4 of the most influential features affecting DFS predictions. The x-axis shows the feature value, and the y-axis shows the SHAP value (ie, the impact on the model's output). Clinical tumor stage showed a strong negative relationship with predicted DFS: as tumor stage increased, the SHAP values shifted sharply downward, indicating a consistent reduction in predicted DFS, aligning with the known prognostic role of tumor invasiveness in bladder cancer. SII demonstrated a nonlinear relationship, showing that patients with lower SII values had better SHAP values, while those with elevated SII showed increasingly negative impacts on DFS. This suggests a threshold effect, where systemic inflammation beyond a certain level contributes to poorer prognosis. The presence of pretreatment hydronephrosis had a negative impact on DFS prediction. Patients with low BMI had negative SHAP values, indicating reduced DFS, while those with moderate BMI experienced mildly negative predictions. At a BMI greater than 28, the SHAP values became positive, suggesting a potential protective effect exerted by higher BMI. Regarding the type of urinary diversion, vesicoileal pouch construction showed positive SHAP values, while other approaches had negative SHAP values.

**Figure 3.** Shapley additive explanations (SHAP) scatterplots for the 5 most significant features influencing disease-free survival predictions (with BMI in kg/m$^2$). CTS: clinical tumor stage (0: cTa, 1: cTis, 2: cT1, 3: cT2, 4: cT3, 5: cT4); PRE-HYDRONEPHROSIS: pretreatment hydronephrosis (0: no, 1: right hydronephrosis, 2: left hydronephrosis, 3: bilateral hydronephrosis); SII: systemic immune-inflammation index; UD: urinary diversion type (0: Bricker ileal conduit, 1: ureterocutaneostomy, 2: vesicoileal pouch).




## OS Prediction

For OS prediction, the CatBoost model achieved an MAE of 17.2 months across the entire patient cohort. When the analysis was restricted to the subgroup of patients who had died (n=156), the prediction error improved to 15.8 months. After filtering features by importance, using a threshold of <0.5, the MAE further improved to 14.6, suggesting that a more compact feature set may improve predictive efficiency without compromising accuracy (Figure 4). This final model was selected for interpretation, as it maintained accuracy while reducing complexity.

**Figure 4.** Progressive improvement in model accuracy. MAE: mean absolute error.

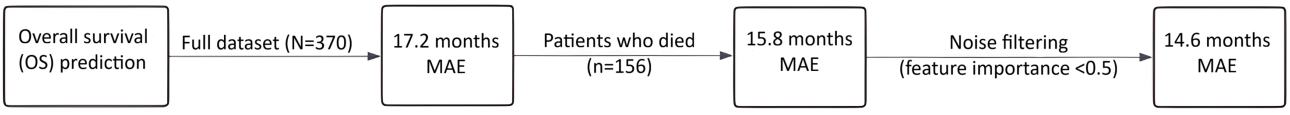



Figure 5 presents the CatBoost feature importance ranking for the best-performing OS prediction model. Tumor classification emerged as the most influential predictor of OS in the final model with a value of approximately 17.5. SII followed closely with a value of approximately 15.5, highlighting the role of systemic inflammation in cancer progression and survival outcomes. The third feature was represented by histological findings (H.E. TURV). Other influencing factors included clinical tumor stage, pretreatment hydronephrosis, type of urinary diversion, BMI, and age.

**Figure 5.** CatBoost feature importance ranking for the best-performing overall survival prediction model. CTS: clinical tumor stage; H.E. TURV: histological examination for transurethral resection of the bladder; LVI: lymphovascular invasion; PRE-HYDRONEPHROSIS: pretreatment hydronephrosis; SII: systemic immune-inflammation index; TC: tumor classification; UD: urinary diversion type.

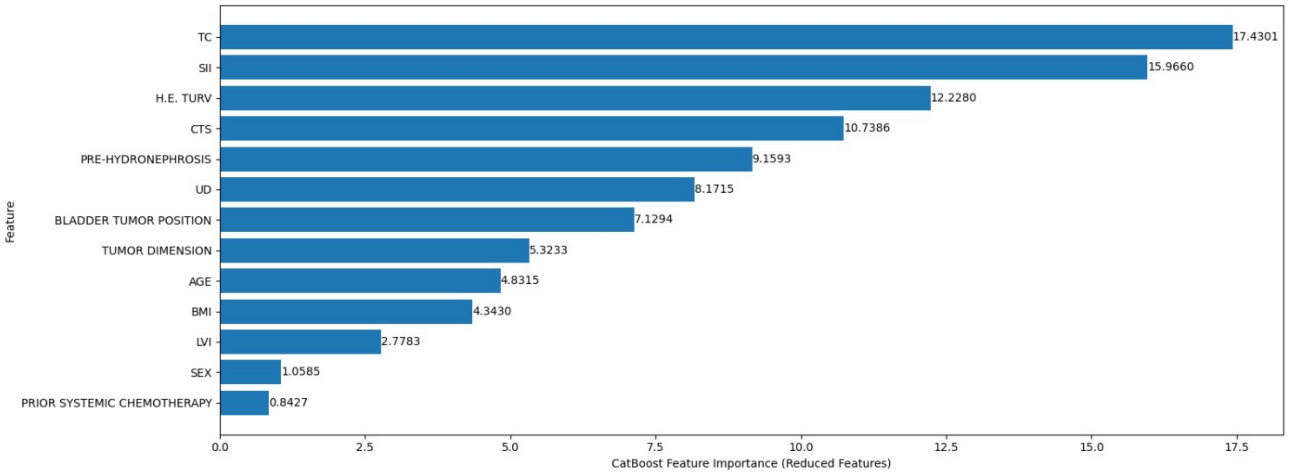



Figure 6 displays the SHAP summary plot for the final OS prediction model. As expected, the most impactful variable was tumor classification, which showed a broad distribution. Pretreatment hydronephrosis and SII exhibited a wide SHAP distribution. Clinical tumor stage and histological findings (H.E. TURV) showed a similar overall effect on prognosis.

**Figure 6.** Shapley additive explanations (SHAP) violin summary plot for the final overall survival prediction model. CTS: clinical tumor stage; H.E. TURV: histological examination for transurethral resection of the bladder; LVI: lymphovascular invasion; PRE-HYDRONEPHROSIS: pretreatment hydronephrosis; SII: systemic immune-inflammation index; TC: tumor classification; UD: urinary diversion type.

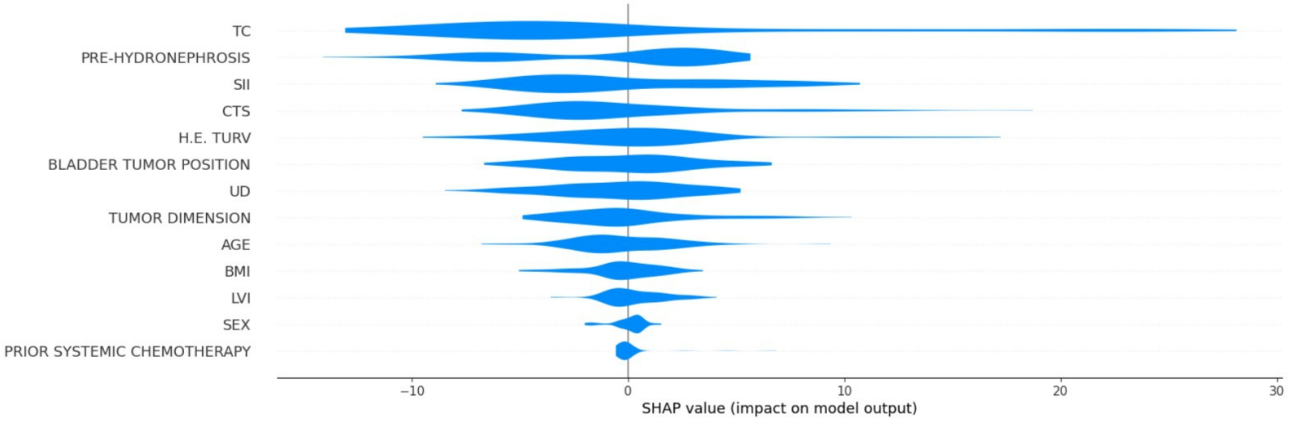
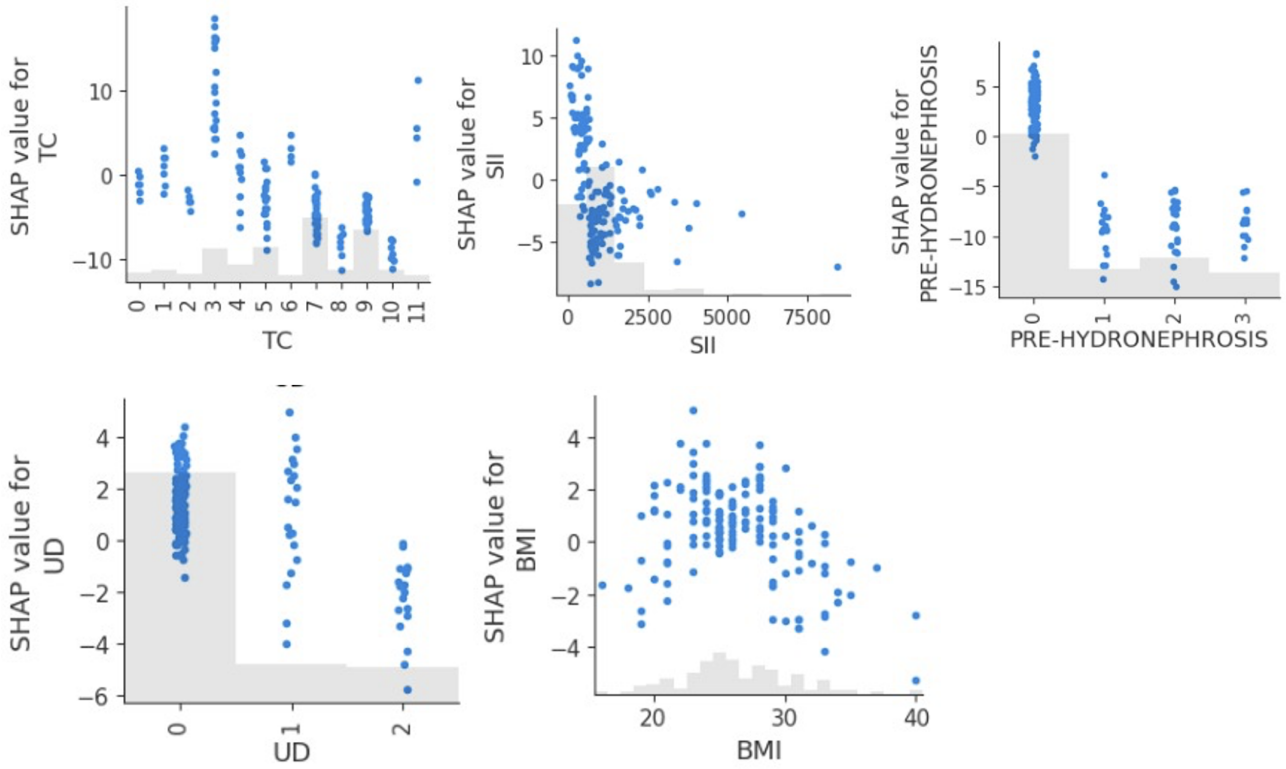



Figure 7 presents the SHAP dependence plots for 5 key features influencing OS predictions. The tumor classification SHAP values indicated that patients with in situ cancers (value 3) had the best OS prediction, which gradually declined as the tumor stage advanced. SII showed a threshold effect, where predictions remained relatively stable up to a value of approximately 1000, then fell sharply, indicating that elevated inflammation is associated with a poor overall outcome. Pretreatment hydronephrosis was strongly linked to reduced predicted OS, where patients with this condition had uniformly negative SHAP values, not influenced by bilaterality. At the same time, BMI demonstrated a nonlinear pattern, where patients with very low BMIs had reduced OS predictions, moderate BMIs were associated with better outcomes, and the SHAP values began to decline again at higher BMI values, suggesting that both underweight and obesity may be associated with increased mortality risk in this population. The type of urinary diversion showed a different impact than that observed in DFS prediction, with vesicoileal pouch construction being associated with a lower OS.

**Figure 7.** Shapley additive explanations (SHAP) scatterplots for the 5 most influential features influencing overall survival prediction (with BMI in kg/m$^2$). PRE-HYDRONEPHROSIS: pretreatment hydronephrosis (0: no, 1: right hydronephrosis, 2: left hydronephrosis, 3: bilateral hydronephrosis); SII: systemic immune-inflammation index; TC: tumor classification (1: T0, 2: Ta, 3: Tis, 4: T1, 5: T2a, 6: T2b, 7: T3a, 8: T3b, 9: T4a, 10: T4b); UD: urinary diversion type (0: Bricker ileal conduit, 1: ureterocutaneostomy, 2: vesicoileal pouch).




## Cause-of-Death Classification

The CatBoostClassifier was trained to predict whether a patient's death was tumor related. Due to class imbalance, only 14 of 78 deaths were cancer related; custom class weights and a reduced decision threshold of 0.12 were applied to maximize recall and minimize false negatives. The final model achieved a recall of 78.6% (Figure 8), correctly identifying 11 of 14 tumor-related deaths. The overall $F_1$-score for the positive class was 0.44, with a precision of 31%. The model prioritizes sensitivity over specificity.

**Figure 8.** Confusion matrix for cause-of-death classification.

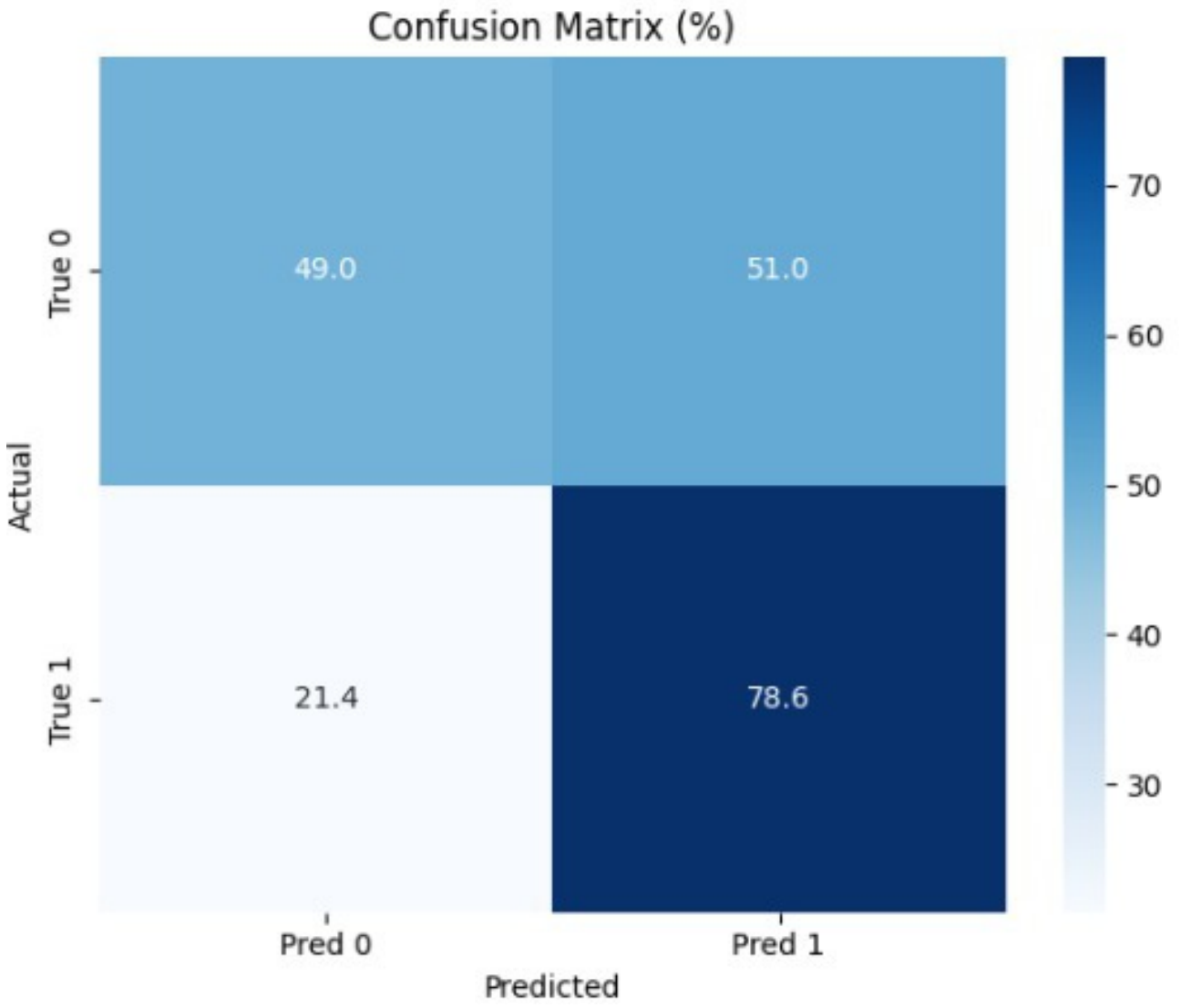
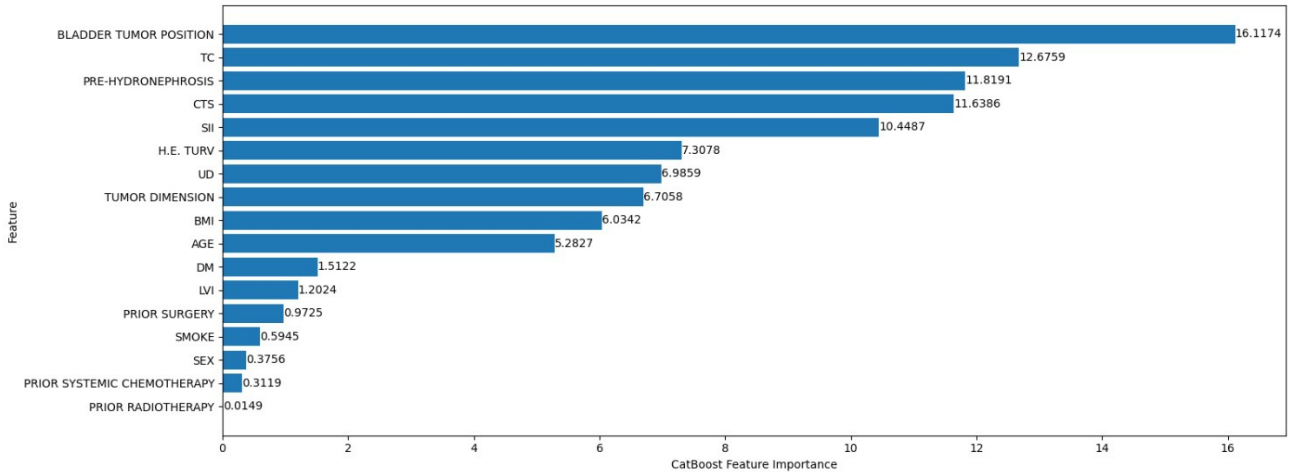



Figure 9 represents the CatBoost feature importance ranking for the death cause classification model. The most influential feature was the anatomical position of the bladder tumor, with a value of approximately 16.5. This was followed by tumor classification and pretreatment hydronephrosis, both indicators of disease severity and progression.

**Figure 9.** CatBoost feature importance ranking for the cause-of-death classification. CTS: clinical tumor stage; DM: diabetes mellitus; H.E. TURV: histological examination for transurethral resection of the bladder; LVI: lymphovascular invasion; PRE-HYDRONEPHROSIS: pretreatment hydronephrosis; SII: systemic immune-inflammation index; TC: tumor classification; UD: urinary diversion type.



Other key features included the clinical tumor stage and SII, with values of approximately 11 and 11.5, respectively.

Figure 10 displays the SHAP summary plot for the tumor-related death classification model. The feature with the widest and most impactful distribution was bladder tumor position, which is addressed in detail in the discussion of Figure 11. SII was also influential, with positive and negative SHAP values. Tumor classification, pretreatment hydronephrosis, and clinical tumor stage generally acted to slightly decrease the predicted risk of tumor-related death for most patients, with little variability in their effect.

**Figure 10.** Shapley additive explanations (SHAP) summary plot for the tumor-related death classification model. CTS: clinical tumor stage; DM: diabetes mellitus; H.E. TURV: histological examination for transurethral resection of the bladder; LVI: lymphovascular invasion; PRE-HYDRONEPHROSIS: pretreatment hydronephrosis; SII: systemic immune-inflammation index; TC: tumor classification; UD: urinary diversion type.

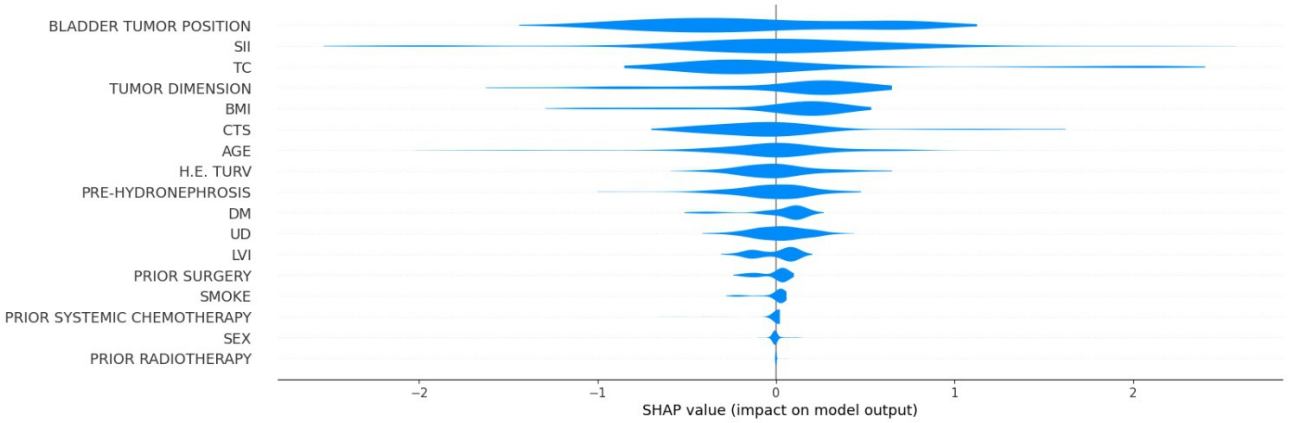



**Figure 11.** Shapley additive explanations (SHAP) scatterplots for the 5 most influential features influencing the tumor-related death classification model (with BMI in kg/m$^2$, AGE in years, and TUMOR DIMENSION in cm; bladder tumor position: 0: intertrigonal zone, 1: right periosteal, 2: left periosteal, 3: dome, 4: posterior wall, 5: right lateral wall, 6: left lateral wall, 7: prostatic urethra, 8: anterior wall, 9: entire bladder, 10: bladder base). SII: systemic immune-inflammation index.

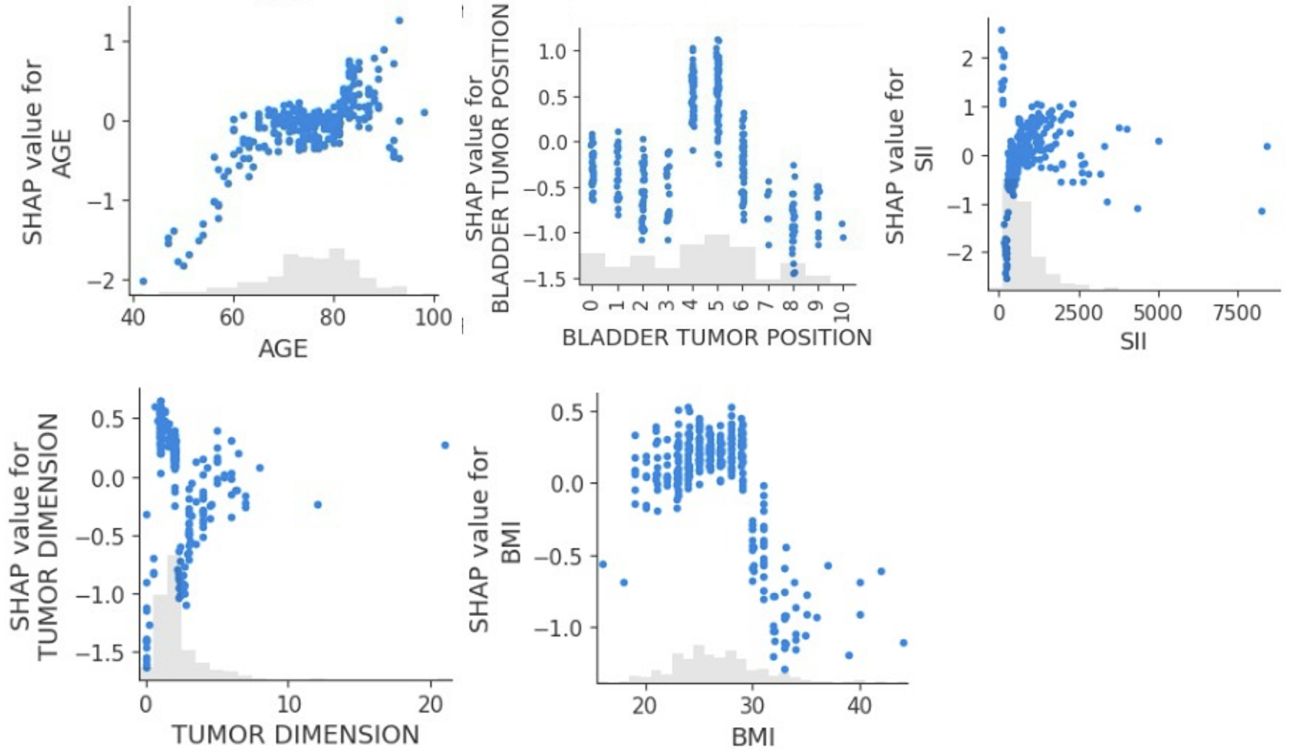



Figure 11 presents the SHAP dependence plots for 5 key features influencing the tumor-related classification model: age, bladder tumor position, SII, tumor dimension, and BMI.

The scatterplot for age exhibited a clear positive trend, showing that as patient age increased, the SHAP value for age also increased, indicating that older age consistently increased the predicted risk of tumor-related death. Bladder tumor position was another influential predictor in the tumor-related death classification model, highlighting that tumors located in the posterior wall and right lateral wall were more likely to be the cause of death. In contrast, tumors located in the anterior wall and at the bladder base or those spreading throughout the entire bladder were less likely to be the cause of death. Increasing SII

and tumor dimension also had a moderate predictive value, respectively increasing and decreasing the probability that the patient's cause of death was cancer. Finally, patients with a higher BMI showed a higher likelihood that the cause of death was cancer.

## Discussion

### Principal Findings

This paper presents the development of a survival prediction model using machine learning approaches for patients with bladder cancer. Our findings demonstrate that modern predictive algorithms show promising accuracy in forecasting DFS, OS, and cause of death. The limited sample size and the paucity of included categories for the analysis suggest that predictive algorithms trained with additional data and variables significantly improve the demonstrated accuracy.

The observation that age showed a positive correlation with survival outcomes is particularly intriguing and seemingly counterintuitive. This "age paradox" has been previously described in bladder cancer and other oncological settings and may reflect a combination of selection bias and underlying tumor biology rather than a true protective effect of age, which can be explained by several factors. Older patients often receive more conservative treatment, which potentially leads to selection bias in surgical candidates. Additionally, younger patients with bladder cancer have been reported to present more frequently with aggressive disease variants, which could account for their relatively poorer outcomes despite younger age [11]. These results align with recent systematic reviews and meta-analyses on radical cystectomy, which consistently report high complication rates and significant variability in postoperative survival across risk profiles [12]. Importantly, as this study is observational, the association between age and survival should be interpreted cautiously and not as a causal relationship.

Clinical tumor stage was the strongest predictor, which aligns with established prognostic factors in bladder cancer [13]. Additionally, inflammatory markers, particularly SII, showed a negative correlation with survival outcomes, supporting recent findings in other malignancies [14]. This relationship likely reflects the complex interplay between systemic inflammation and cancer progression, where elevated SII indicates a protumoral inflammatory state [15]. Our findings on systemic inflammatory indices are consistent with recent data indicating that platelet-to-lymphocyte ratio, systemic inflammation response index, pan-immune-inflammation value, SII, and neutrophil-to-lymphocyte ratio are associated with adverse outcomes in non–muscle-invasive bladder cancer [16].

SHAP analysis revealed a clear, monotonic decline in predicted survival with advancing clinical tumor stage, reinforcing its primary prognostic role in both DFS and OS. SII, by contrast, demonstrated a threshold effect where values above approximately 1000 were associated with a sharp drop in predicted DFS, suggesting a nonlinear relationship between systemic inflammation and patient outcomes. The observed association between urinary diversion type and survival outcomes should be considered exploratory. Previous studies had found that orthotopic neobladder reconstruction had a protective effect against urethral recurrence in male patients undergoing radical cystectomy for bladder cancer [17]. While this effect was not observed in our dataset, we found a positive association between vesicoileal pouch construction and improved survival. This association may reflect both patient selection and potential physiological advantages of this diversion type; however, given the limited number of patients within each diversion subgroup, it should be interpreted cautiously. Confounding factors such as surgical expertise and patient characteristics, which were not accounted for in the present study, may have influenced these findings. BMI showed a similar intriguing relationship: patients with an unhealthy BMI, either high or low, showed poorer outcomes; this may be related both to tumor characteristics and the surgical approach being limited in terms of radicality. This U-shaped association between BMI and survival was consistently observed across both DFS and OS outcomes, with moderate BMI ranges correlated with more favorable SHAP values. The findings support a metabolic vulnerability in patients with underweight as well as obesity, which may influence recovery or treatment tolerance.

Our machine learning models achieved prediction accuracies comparable to those reported in previous studies. The accuracy in cause-of-death prediction, although modest, represents an encouraging level, given the limited resources and the paucity of categories considered for the analysis, particularly when compared with studies published a few years ago that used significantly larger samples yet achieved marginally higher accuracy in mortality and recurrence prediction [18]. A recently published systematic review investigating machine learning algorithms for bladder cancer cystectomy outcomes found that most of the algorithms did not exceed 70% accuracy and, in some cases, performed with approximately 60% accuracy [19]. The integration of SII into predictive models represents an auspicious direction. As a low-cost, readily available biomarker, SII could enhance current prognostic tools without adding significant complexity or cost to patient evaluation [14].

A notable limitation of our study is the relatively high MAE in survival predictions. These MAE values render the algorithm unsuitable for precise individual patient counseling or treatment planning where accurate timing is critical, such as in emergency settings or for patients exhibiting postoperative complications [20]. However, this level of accuracy remains acceptable for clinical trial patient stratification and allocation, particularly in trials where broad risk categories rather than precise survival estimates are needed for randomization. Such applications include balancing treatment arms in clinical trials by identifying comparable risk groups or supporting enrollment decisions in competing risk analysis, where precise timing is less critical than overall risk assessment [21].

### Limitations and Reproducibility

This study is subject to some limitations when interpreting the results. The relatively limited dataset size (N=370 initially; reduced to 312 - 347 for specific analyses) inherently constrains the generalizability and robustness of the developed models. While machine learning algorithms such as CatBoost are robust

on smaller datasets, their predictive power can be substantially enhanced with larger cohorts.

Secondly, the monocentric nature of the data collection, originating solely from Fondazione Policlinico Universitario Agostino Gemelli IRCCS in Rome, Italy, introduces a potential for selection bias and limits external validity. Patient characteristics, treatment protocols, and population demographics can vary significantly across institutions and geographical regions. The findings from this study may not be directly transferable to other clinical settings without further validation on diverse, external datasets.

Thirdly, while rigorous data cleaning was performed, the inherent human factors associated with retrospective data extraction from medical records cannot be eliminated.

## Conclusion

Our study demonstrates the potential utility of machine learning approaches in predicting bladder cancer outcomes following cystectomy. While the achieved accuracy levels are modest, they align with current literature benchmarks and provide a foundation for future development. The identification of clinical tumor stage as the primary predictor, along with the consistent negative correlation of SII with survival outcomes, validates these parameters as valuable prognostic indicators. In particular, the SHAP analysis revealed a monotonic decline in predicted DFS and OS with advancing clinical tumor stage, reaffirming its role in risk stratification. On the other hand, SII exhibited a threshold effect, where values above approximately 1000 were associated with a rapid drop in predicted survival, reinforcing the adverse prognostic impact of systemic inflammation. The current model's performance, though not suitable for precise individual prognostication, shows particular promise for clinical trial stratification and cohort allocation. Future studies with larger datasets and additional predictive variables may enhance the model's accuracy and broaden its clinical applications. Integrating readily available biomarkers, such as SII, represents a cost-effective approach to improving prognostic tools. These findings contribute to the growing body of evidence supporting the role of machine learning in oncological decision-making while acknowledging the need for continued refinement and validation in larger cohorts.

## Funding

No external financial support or grants were received from any public, commercial, or not-for-profit entities for the research, authorship, or publication of this article.

## Data Availability

To ensure the reproducibility of our results and facilitate further research, all code used for this analysis has been made publicly available on Figshare [10].

## Authors' Contributions

FAC, BR, and PR elaborated on the first manuscript concept. AN and FAC performed the statistical analysis. FAC, VDV, and PR wrote the article. MS, NF, and GM reviewed and approved the final manuscript.

## Conflicts of Interest

None declared.

## References

1. Anagnostou T, Remzi M, Lykourinas M, Djavan B. Artificial neural networks for decision-making in urologic oncology. Eur Urol 2003 Jun;43(6):596-603. [doi: 10.1016/s0302-2838(03)00133-7] [Medline: 12767358]
2. Schwarzer G, Schumacher M. Artificial neural networks for diagnosis and prognosis in prostate cancer. Semin Urol Oncol 2002 May;20(2):89-95. [doi: 10.1053/suro.2002.32492] [Medline: 12012294]
3. Sargent DJ. Comparison of artificial neural networks with other statistical approaches: results from medical data sets. Cancer 2001 Apr 15;91(8 Suppl):1636-1642. [doi: 10.1002/1097-0142(20010415)91:8+<1636::aid-cncr1176>3.0.co;2-d] [Medline: 11309761]
4. Zi H, Liu MY, Luo LS, et al. Global burden of benign prostatic hyperplasia, urinary tract infections, urolithiasis, bladder cancer, kidney cancer, and prostate cancer from 1990 to 2021. Mil Med Res 2024 Sep 18;11(1):64. [doi: 10.1186/s40779-024-00569-w] [Medline: 39294748]
5. Abbod MF, Catto JWF, Linkens DA, Hamdy FC. Application of artificial intelligence to the management of urological cancer. J Urol 2007 Oct;178(4 Pt 1):1150-1156. [doi: 10.1016/j.juro.2007.05.122] [Medline: 17698099]
6. Di Sarno L, Caroselli A, Tonin G, et al. Artificial intelligence in pediatric emergency medicine: applications, challenges, and future perspectives. Biomedicines 2024 May 30;12(6):1220. [doi: 10.3390/biomedicines12061220] [Medline: 38927427]
7. Mazzotta AD, Burti E, Causio FA, et al. Machine learning approaches for the prediction of postoperative major complications in patients undergoing surgery for bowel obstruction. J Pers Med 2024 Oct 8;14(10):1043. [doi: 10.3390/jpm14101043] [Medline: 39452550]
8. Tokuc E, Eksi M, Kayar R, et al. Inflammation indexes and machine-learning algorithm in predicting urethroplasty success. Investig Clin Urol 2024 May;65(3):240-247. [doi: 10.4111/icu.20230302] [Medline: 38714514]

9. Yin M, Joshi M, Meijer RP, et al. Neoadjuvant chemotherapy for muscle-invasive bladder cancer: a systematic review and two-step meta-analysis. Oncologist 2016 Jun;21(6):708-715. [doi: 10.1634/theoncologist.2015-0440] [Medline: 27053504]
10. Bladder cancer CatBoost analysis. Figshare. URL: https://figshare.com/s/ec2419a33ad7ad05f412 [accessed 2026-01-03]
11. Withrow DR, Nicholson BD, Morris EJA, Wong ML, Pilleron S. Age-related differences in cancer relative survival in the United States: a SEER-18 analysis. Int J Cancer 2023 Jun 1;152(11):2283-2291. [doi: 10.1002/ijc.34463] [Medline: 36752633]
12. Sacco E, Campetella M, Marino F, et al. Preoperative risk factors for failure after fixed sling implantation for postprostatectomy stress urinary incontinence (FORESEE): a systematic review and meta-analysis. Eur Urol Open Sci 2025 Jun;77:58-77. [doi: 10.1016/j.euros.2025.05.012] [Medline: 40546391]
13. Kardoust Parizi M, Enikeev D, Glybochko PV, et al. Prognostic value of T1 substaging on oncological outcomes in patients with non-muscle-invasive bladder urothelial carcinoma: a systematic literature review and meta-analysis. World J Urol 2020 Jun;38(6):1437-1449. [doi: 10.1007/s00345-019-02936-y] [Medline: 31493109]
14. Kou J, Huang J, Li J, Wu Z, Ni L. Systemic immune-inflammation index predicts prognosis and responsiveness to immunotherapy in cancer patients: a systematic review and meta-analysis. Clin Exp Med 2023 Nov;23(7):3895-3905. [doi: 10.1007/s10238-023-01035-y] [Medline: 36966477]
15. Wang Q, Zhu D. The prognostic value of systemic immune-inflammation index (SII) in patients after radical operation for carcinoma of stomach in gastric cancer. J Gastrointest Oncol 2019 Oct;10(5):965-978. [doi: 10.21037/jgo.2019.05.03] [Medline: 31602335]
16. Bizzarri FP, Campetella M, Russo P, et al. Prognostic value of PLR, SIRI, PIV, SII, and NLR in non-muscle invasive bladder cancer: can inflammatory factors influence pathogenesis and outcomes? Cancers (Basel) 2025 Jun 28;17(13):2189. [doi: 10.3390/cancers17132189] [Medline: 40647487]
17. Yu J, Lee CU, Chung JH, et al. Impact of urinary diversion type on urethral recurrence following radical cystectomy for bladder cancer: propensity score matched and weighted analyses of retrospective cohort. Int J Surg 2024 Feb 1;110(2):700-708. [doi: 10.1097/JS9.0000000000000904] [Medline: 38000052]
18. Hasnain Z, Mason J, Gill K, et al. Machine learning models for predicting post-cystectomy recurrence and survival in bladder cancer patients. PLoS One 2019;14(2):e0210976. [doi: 10.1371/journal.pone.0210976] [Medline: 30785915]
19. Liu YS, Thaliffdeen R, Han S, Park C. Use of machine learning to predict bladder cancer survival outcomes: a systematic literature review. Expert Rev Pharmacoecon Outcomes Res 2023;23(7):761-771. [doi: 10.1080/14737167.2023.2224963] [Medline: 37306511]
20. Qi S, Kumar N, Farrokh M, et al. An effective meaningful way to evaluate survival models. arXiv. Preprint posted online on Jun 1, 2023. [doi: 10.48550/arXiv.2306.01196]
21. Hassan N, Slight R, Morgan G, et al. Road map for clinicians to develop and evaluate AI predictive models to inform clinical decision-making. BMJ Health Care Inform 2023 Aug;30(1):e100784. [doi: 10.1136/bmjhci-2023-100784] [Medline: 37558245]


## Abbreviations

**AI:** artificial intelligence
**DFS :** disease-free survival
**MAE:** mean absolute error
**OS:** overall survival
**SHAP:** Shapley additive explanations
**SII:** systemic immune-inflammation index
**TRIPOD:** Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis


*Edited by N Rohatgi; submitted 28.Oct.2025; peer-reviewed by E Tokuc, FP Bizzarri; revised version received 09.Jan.2026; accepted 13.Jan.2026; published 19.Feb.2026.*

*Please cite as:*
*Causio FA, De Vita V, Nappi A, Sawaya M, Rocco B, Foschi N, Maioriello G, Russo P*
*Survival Prediction in Patients With Bladder Cancer Undergoing Radical Cystectomy Using a Machine Learning Algorithm: Retrospective Single-Center Study*
*JMIR Perioper Med 2026;9:e86666*
*URL: https://periop.jmir.org/2026/1/e86666*
*doi:10.2196/86666*

# Physician Perspectives on ChatGPT-4o as a Patient Resource for Abdominal Cancer Surgeries: Cross-Sectional Survey

Christina V Lindsay[1,2], BS; Devika A Shenoy[1,2], BS; Allison N Martin[3], MPH, MD; Christie L Clipper[2], DHA; Kevin N Shah[3], MD; Michael E Lidsky[3], MD; Daniel P Nussbaum[3], MD; Ralph Snyderman[2,4], MD

[1]School of Medicine, Duke University, 8 Searle Center Dr, Durham, NC, United States

[2]Center for Personalized Health Care, Duke University, Durham, NC, United States

[3]Department of Surgery, Duke University, Durham, NC, United States

[4]Department of Medicine, Duke University, Durham, NC, United States

**Corresponding Author:**
Christina V Lindsay, BS
School of Medicine, Duke University, 8 Searle Center Dr, Durham, NC, United States

## Abstract

**Background:** Artificial intelligence (AI) models are being increasingly integrated into clinical care. Moreover, the availability of publicly accessible AI resources makes them attractive to patients seeking clinical information. Little is known regarding the use of large language models as patient resources for navigating major cancer diagnoses.

**Objective:** This study aimed to evaluate the content, readability, and safety of ChatGPT (OpenAI; GPT-4o)-generated responses to common perioperative queries about hepatic, pancreatic, and colon cancers.

**Methods:** A 28-question survey was developed based on frequently asked surgical questions for select malignancies. Surgical oncologists rated ChatGPT-4o-generated responses on a 5-point Likert scale for accuracy, quality, and tangibility. Readability was assessed using the Flesch-Kincaid Reading Grade Level (FKRGL) and Flesch Reading Ease (FRE). Respondents provided free-text comments and reported their comfort with patients using ChatGPT. Survey completion implied consent.

**Results:** A total of 7 attending surgical oncologists with a median of 7 (IQR 4-13) years in practice completed the survey. Responses received mean scores of 3.5/5 (SD 0.28) for quality, 3.6/5 (SD 0.34) for accuracy, and 3.6/5 (SD 0.29) for tangibility. The responses had a median FKRGL score of 14.6 (IQR 13.3-15.6) and FRE score of 29.4 (IQR 20.5-36.3). On a post hoc analysis for select questions, the median FKRGL was 15.6 (IQR 14.4-16.7), decreasing to 7.1 (IQR 6.1-8.3) and 14.5 (IQR 13.2-15.4) with prompting and rephrasing, and the median FRE was 18.1 (IQR 14.6-24.7), increasing to 73.8 (IQR 66.6-79.3) and 32.0 (IQR 27.0-37.7) with prompting and rephrasing. Numerous inaccuracies and content gaps were reported, and approximately 43% (3/7) of providers did not report feeling “comfortable” in having patients consult publicly available AI for medical information.

**Conclusions:** This study provides cautionary, yet optimistic, findings regarding the value of publicly accessible ChatGPT as a patient resource for abdominal malignancies. Providers should be prepared to effectively counsel patients to identify their educational attainment level when using ChatGPT to mitigate readability challenges.

***(JMIR Perioper Med 2026;9:e81374)*** doi:10.2196/81374

**KEYWORDS**

patient education; health literacy; generative artificial intelligence; surgical oncology; perioperative care

## Introduction

In recent years, artificial intelligence (AI) has promised to reshape medicine. Chatbots such as ChatGPT (OpenAI) [1], DeepAI, and Google Gemini use large language models (LLMs), a popular form of AI. These models are trained upon large datasets to generate answers [2]. Recent LLM improvement in reasoning has been noted to reflect human-level cognition [3]. Furthermore, studies have examined LLM function in the health care sector. LLMs have been found to pass United States Medical Licensing Examinations [4,5] and medical subspecialty exams [6-8] and to provide successful clinical reasoning and diagnoses [9]. Moreover, ChatGPT has the potential to supersede other search engines in answering patient health–related questions by providing more comprehensive and specific answers [10,11].

Although AI has been found to augment medical practice, its use as a resource by patients is not well understood. Patients have long reported turning to the internet for clinical advice [12]. Studies evaluating responses from common search engines to frequently asked general surgery questions have typically found the quality to range from fair to good but found that the readability level often exceeded the recommended level for the general population [13]. More recently, patients have turned to

ChatGPT for clinical questions; a study conducted in Australia found that approximately 9.9% of Australian adults asked ChatGPT medical questions within the first half of 2024 [14]. Following the rapid rise of publicly accessible proprietary LLM chatbots and the lack of peer-reviewed output within these learning models, recent work across specialties, including oncology, gastroenterology, otolaryngology, and surgery, has sought to evaluate LLM-generated responses to questions commonly asked by patients [15-21]. The reported overall quality of generated responses varies across fields and is impacted by the type of LLM used [20]. Furthermore, prior research has suggested that, as with "Dr Google" and other popular search engines [13], the readability of LLM-generated responses may serve as a key limitation of using LLMs such as ChatGPT as a patient resource [22]. Additionally, ChatGPT answers are limited in consistency [23], generating similar but nonidentical responses.

Gastrointestinal malignancies, including pancreatic, colorectal, hepatic, stomach, and esophageal malignancies, account for over one-quarter of cancer incidence globally and are steadily increasing. By 2040, the global number of gastrointestinal cancer deaths is projected to increase by over 70% to 5.6 million [24]. Given the significant disease burden of gastrointestinal malignancies and related therapies, it is essential to properly evaluate pertinent patient resources to better inform patients, many of whom will be accessing these resources independently. To date, few studies have examined the use of publicly accessible proprietary LLMs as a perioperative resource for patients with abdominal malignancies. The aim of this study was to evaluate the content and readability of LLM-generated responses to common patient queries for hepatic, pancreatic, and colon cancers.

## Methods

### Ethical Considerations

This study was submitted to the Duke University Institutional Review Board for review and was determined to be exempt (Pro00116649). This study involved surgeon-participants who evaluated GPT-generated responses to frequently asked questions. No patient data were used. To maintain participant confidentiality, all data were analyzed in aggregate. All GPT inquiries and survey questions were asked in English. Consent from participants was implied through voluntary completion of the survey. No compensation was provided to participants.

### Question Development

Preliminary questions were developed by CVL and DAS. Questions for this study were developed by sourcing frequently asked questions about colon, liver, and pancreatic cancers from 7 hospital patient information and nonprofit cancer foundation websites [25-31]. This methodology was used in an earlier study examining LLMs as a tool for patient education in lung cancer surgery [32]. Questions on general disease information, including signs and symptoms, staging and treatment options, surgical eligibility, and operative risks, were formulated for colon, hepatic, and pancreatic cancers using identical language for each condition. Standardized language was used to determine the suitability of LLMs for delivering useful abdominal cancer education as applied to each of the conditions. Additional questions were created to address common patient concerns related to postoperative recovery and potential adverse outcomes following abdominal cancer surgery.

### Question Piloting

The preliminary survey questions were initially evaluated for relevance and alignment with patient phrasing through subjective assessment by 2 general surgery residents. Residents were prompted to assess the frequency with which the proposed questions were encountered in practice to evaluate the survey questions based on clinical relevance. Residents were also prompted to evaluate the survey questions based on alignment with patient phrasing, to suggest phrasing revisions for items that received a Likert score ≤3 on a 5-point scale, and to propose additional relevant questions not addressed by the survey. At this stage, 8 questions were removed and 4 were added per resident feedback. Additional questions were adjusted accordingly. After piloting scenarios with residents, all questions were run sequentially through a publicly available, proprietary version of ChatGPT (GPT-4o; released on May 13, 2024) [1,33], on March 9, 2025, in Durham, United States. ChatGPT-4o is based on a proprietary GPT-4-class pretrained base LLM that has been instruction-tuned for conversational use. No additional model fine-tuning or retraining was performed through this study. ChatGPT was prompted to answer in paragraph form without additional contextual information. As in prior studies, a new chat entry was posed for each question [32].

A Qualtrics survey was formulated with the final 28 questions and LLM responses. This survey was piloted by our surgeon expert, ANM, who provided final revisions for question phrasing. Table 1 lists the finalized questions prompted into ChatGPT. Revised questions were newly run through ChatGPT, and the Qualtrics survey was adjusted accordingly.

**Table .** GPT queries: abdominal cancer frequently asked questions and common postoperative complications.

| Domain | Question |
|---|---|
| Signs and symptoms | Q1: What are the signs and symptoms of pancreatic cancer? |
| | Q2: What are the signs and symptoms of colon cancer? |
| | Q3: What are the signs and symptoms of liver cancer? |
| Stages and treatment | Q4: What are the different stages and treatments for pancreatic cancer? |
| | Q5: What are the different stages and treatments for colon cancer? |
| | Q6: What are the different stages and treatments for liver cancer? |
| Surgery eligibility | Q7: Who is appropriate for surgery for pancreatic cancer? |
| | Q8: Who is appropriate for surgery for colon cancer? |
| | Q9: Who is appropriate for surgery for liver cancer? |
| Surgery risks | Q13: What are the risks of surgery to remove my pancreatic cancer? |
| | Q14: What are the risks of surgery to remove my colon cancer? |
| | Q15: What are the risks of surgery to remove my liver cancer? |
| General postoperative recovery | Q10: How long is the recovery from pancreatic cancer surgery? |
| | Q11: How long is the recovery from colon cancer surgery? |
| | Q12: How long is the recovery from liver cancer surgery? |
| | Q16: Will I need an ostomy after surgery to remove my colon cancer? |
| | Q17: How long will I be in the hospital after surgery for cancer in my belly? |
| | Q18: How long after surgery for cancer in my belly can I exercise? |
| | Q19: How long will it take to recover from surgery for cancer in my belly? |
| | Q20: Should I stay close to the hospital in a hotel or Airbnb after I'm discharged from surgery for cancer in my belly? |
| | Q21: How long after surgery for cancer in my belly can I do chores around the house? |
| Adverse outcomes | Q22: I just had surgery for cancer in my belly, and my incision is painful. What do I do? |
| | Q23: I just had surgery for cancer in my belly, and I am still in some pain. Is there anything else I can take for the pain? |
| | Q24: I just had surgery for cancer in my belly, and my incision is starting to hurt more and looks slightly open. What do I do? |
| | Q25: I just had surgery for cancer in my belly, and the incision is warm to the touch and draining a yellowish fluid. What do I do? |
| | Q26: I just had surgery for cancer in my belly. It hurts when I breathe, and I have a new cough. What do I do? |
| | Q27: I just had surgery for cancer in my belly. It now burns when I pee. What do I do? |
| | Q28: I am about to have surgery for cancer in my belly. How can I prevent an infection after? |

## Outcomes or Data Collection and Variables

An anonymous survey was disseminated to surgical faculty at a single institution using Qualtrics, a secure, web-based survey platform. Eligible participants were board-certified surgeons who had completed fellowship training in surgical oncology or colorectal surgery and were actively practicing at the time of the study. Surgeons were identified through publicly available web-based colorectal surgery and surgical oncology faculty rosters and were invited to participate via an email containing the anonymous Qualtrics link. The Qualtrics platform is commonly used in academic research, as it permits investigators to design surveys, test them for accessibility and functionality, distribute them electronically as a web link or QR code, and export result reports. The final list of questions and ChatGPT-4o–generated responses graded by surgeons is included in Multimedia Appendix 1. On the Qualtrics platform, prior to initiating the survey, surgeons were instructed to grade responses for accuracy, quality, and tangibility on a 5-point Likert scale (1="poor," 5="excellent"). The survey defined accuracy as the medical or social correctness of a response, quality as the extent to which a response is well-written and comprehensive, and tangibility as the degree to which the response provides actionable guidance.

After evaluating ChatGPT responses, surgeon respondents were prompted to self-report demographics, including age, sex, years as a practicing surgeon post training, AI frequency in practice, and prior experience with AI. Furthermore, a free response section permitted surgeons to share thoughts or concerns.

Respondents were assessed for comfort with patient-AI use through the question, “If a patient informed you they are using publicly available AI (eg, ChatGPT) for health information, how comfortable would you be with encouraging them to use AI following this survey?” Available answer choices included “very uncomfortable,” “uncomfortable,” “neither comfortable nor uncomfortable,” “comfortable,” and “very comfortable.”

LLM-generated responses were separately graded for readability using the Flesch-Kincaid Reading Grade Level (FKRGL) and Flesch Reading Ease (FRE) formulas through the Readability Statistics tool in Microsoft Word Version 16.105.2 [34]. FRE and FKRGL formulas calculate readability based on the average sentence and word length of a text. The FKRGL scale assesses approximate grade level of a text, with an FKRGL score of 5 corresponding to a US 5th-grade reading level. The FRE scale measures readability from 0, unreadable text, to 100, very easily readable text. Both scales were selected as they are validated tools for grading text readability, and they are commonly used by professionals to evaluate the readability of patient-directed health care information [34].

A post hoc analysis was performed to assess FKRGL, FRE scores, and content similarity for GPT responses under 3 prompting conditions: version 1 represented the response to the original question provided for reference; version 2 consisted of responses to the original question preceded by a prompt to “Answer at a 5th-grade level;” and version 3 comprised responses to questions that were reworded to a 5th-grade reading level by ChatGPT-4o prior to response generation. Four questions were selected for post hoc analysis to provide a focused analysis of question-phrasing and prompting on readability. Questions were selected based on having the highest original FKRGL score within 4 different domains and to ensure representation of each malignancy type. Content similarity was graded by 2 independent graders, CVL and DAS, using a 5-point Likert scale (1=not similar, 5=very similar).

### Statistical Analysis

Respondent answers were collected and analyzed in aggregate. Descriptive statistics for categorical variables were reported as frequencies with percentages; continuous variables were reported as mean with SD or median with IQR, where appropriate. Analyses were calculated using Microsoft Excel Version 16.95.4; formulas used included =MEDIAN() for median, =AVERAGE() for mean, =STDEV() for SD, and =QUARTILE.INC() to derive the IQR.

## Results

### Overview

Of the 12 eligible surgeons contacted, 7 responded, resulting in a survey response rate of 58.3%. All respondents were academic surgeons at a single institution. The median reported respondent categorical age range was 35 - 44 years. Most survey respondents were male (4/7, 57.1%). Respondents had practiced surgery for a median of 7 (IQR 4 - 13) years post training. When assessed for frequency of AI use, 1 respondent reported using AI “daily,” 2 reported using AI “weekly,” 3 reported using AI “monthly,” and 1 reported using AI “almost never.”

### Quality of LLM Responses

When asked to evaluate the quality of responses, experts consistently rated answers between “good” and “very good” to “excellent,” with an aggregate mean response rating of 3.54 (SD 0.28). Across all domains (Table 2), questions related to staging and treatment consistently performed worse, receiving an average rating of 3.33 (SD 0.30), while questions about adverse outcomes tended to perform best, receiving an average rating of 3.73 (SD 0.27). Table 2 indicates the median and IQR of respondent grade for each question, with quality scores ranging from 3.00 to 4.50 and IQR ranging from 2.50-3.50 to 3.25-5.00. The question indicating postoperative urinary tract infection (UTI) received the highest median quality score of 4.50 (3.25 - 5.00), between “very good” and “excellent.”

**Table .** Quality, accuracy, and tangibility scores for GPT-generated responses[a].

| Domain and question ID | Quality score, median (IQR) | Accuracy score, median (IQR) | Tangibility score, median (IQR) |
|---|---|---|---|
| Signs and symptoms | | | |
| Q1 | 4 (3.5 - 4) | 4 (4 - 4.5) | 4 (4 - 4.5) |
| Q2 | 3 (3-4) | 4 (3.5 - 4.5) | 4 (3-4) |
| Q3 | 3 (2.5 - 4) | 3 (2.5 - 4) | 3 (3 - 3.5) |
| Stages and treatment | | | |
| Q4 | 3 (2.5 - 4.5) | 4 (2.5 - 4.5) | 4 (2.5 - 4.5) |
| Q5 | 4 (3-4) | 4 (3-4) | 4 (3.5 - 4) |
| Q6 | 3 (2.5 - 3.5) | 3 (2.5 - 3.5) | 3 (2.5 - 3.5) |
| Surgery eligibility | | | |
| Q7 | 4 (2.5 - 4) | 4 (2.5 - 4) | 4 (3.5 - 4) |
| Q8 | 4 (3.5 - 4) | 4 (3.5 - 4) | 4 (3.5 - 4) |
| Q9 | 4 (4-4) | 4 (3.5 - 4) | 4 (4-4) |
| General postoperative recovery | | | |
| Q10 | 4 (3-4) | 4 (2.5 - 4) | 4 (3-4) |
| Q11 | 3 (2.5 - 4) | 3 (2.5 - 3.5) | 3 (2.5 - 3.5) |
| Q12 | 4 (2.5 - 4) | 3 (2.5 - 3.5) | 3 (2.5 - 4) |
| Surgery risks | | | |
| Q13 | 4 (3-4) | 4 (3-4) | 3 (3-4) |
| Q14 | 4 (3.5 - 4) | 4 (3.5 - 4) | 4 (3.5 - 4) |
| Q15 | 3 (3 - 3.5) | 3 (3 - 3.5) | 3 (3 - 3.5) |
| General postoperative recovery | | | |
| Q16 | 4 (3-4) | 4 (3.5 - 4) | 4 (3.5 - 4) |
| Q17 | 3 (2.5 - 4) | 4 (3.5 - 4) | 4 (2.5 - 4) |
| Q18 | 4 (3.5 - 4.5) | 4 (3.5 - 4.5) | 4 (3.5 - 4) |
| Q19 | 4 (3.5 - 4) | 4 (3.5 - 4) | 4 (3-4) |
| Q20 | 4 (3.5 - 4) | 4 (3-4) | 4 (3-4) |
| Q21 | 4 (3-4) | 4 (3-4) | 4 (3 - 4.5) |
| Adverse outcomes | | | |
| Q22 | 3.5 (3-4) | 3.5 (3-4) | 3.5 (3-4) |
| Q23 | 3.5 (3-4) | 4 (3.25 - 4) | 4 (3.25 - 4) |
| Q24 | 4 (3.25 - 4.75) | 4.5 (3.25 - 5) | 4.5 (3.25 - 5) |
| Q25 | 4 (3.25 - 4) | 4 (3.25 - 4) | 4 (3.25 - 4) |
| Q26 | 4 (3.25 - 4.75) | 4.5 (3.25 - 5) | 4.5 (3.25 - 5) |
| Q27 | 4.5 (3.25 - 5) | 4.5 (3.25 - 5) | 4.5 (3.25 - 5) |
| Q28 | 4 (4-4) | 4 (3.25 - 4.75) | 4.5 (3.25 - 5) |

[a]Accuracy is defined as how "medically or socially accurate" a response is; quality as how "well-written and comprehensive" a response is; and tangibility as how "actionable" a response is.

## Accuracy of LLM Responses

Similar to quality, when asked to evaluate the accuracy of responses, experts generally rated responses between "good" and "very good" to "excellent," with an aggregate mean response rating of 3.57 (SD 0.34). Across all domains, questions related to staging and treatment performed the worst, receiving an average rating of 3.29 (SD 0.38). Conversely, questions about adverse outcomes consistently performed best, receiving an average rating of 3.83 (SD 0.24). Median response accuracy ratings ranged from 3.00 to 4.50 with IQRs ranging from 2.50-3.50 to 3.25-5.00. Questions regarding postoperative wound

dehiscence (Q24), pulmonary embolism (Q26), and UTI management (Q27) received the highest median accuracy grading of 4.50 (IQR 3.25-5.00).

### Tangibility of LLM Responses

When asked to evaluate the tangibility, or how "actionable" a response was, experts likewise consistently rated responses between "good" and "very good" to "excellent," with an aggregate mean response rating of 3.62 (SD 0.29). Across all domains, questions pertaining to staging and treatment performed the worst, receiving the lowest mean tangibility score of 3.47 (SD 0.30), while questions about adverse outcomes performed best, receiving an average rating of 3.86 (SD 0.28). The median response ratings ranged from 3.00 to 4.50 with IQR scores ranging from 2.50-3.50 to 3.25-5.00. Questions regarding postoperative wound dehiscence (Q24), pulmonary embolism (Q26), UTI management (Q27), and infection prevention (Q28) received the highest median tangibility grading of 4.50 (IQR 3.25-5.00).

### Readability of LLM Responses

When assessing readability (Table 3), ChatGPT-4o–generated responses read at an average FKRGL of 14.51 (SD 1.86), requiring some level of college education for adequate comprehension. Response FKRGL scores ranged from 10.8 to 18.1. A question regarding wound dehiscence received the lowest grade score, 10.8, while a question regarding pancreatic cancer surgery candidacy received the highest score, 18.1. The mean FRE score of ChatGPT-4o–generated responses was 28.8 (SD 9.87), corresponding to a college graduate reading level and indicating low readability. Response FRE scores ranged from 11.7 to 48.0. A question regarding colon cancer surgery recovery had the worst readability, with an FRE score of 11.7. As with FKRGL grading, a question regarding wound dehiscence had the highest ease of readability with an FRE score of 48.0.

**Table .** Readability of GPT-generated responses. Readability is represented as FKRGL[a] and FRE[b] scores.

| ID and question | FKRGL score (US-grade reading level) | FRE score | Estimated FRE US-grade level [34] |
|---|---|---|---|
| Q1: What are the signs and symptoms of pancreatic cancer? | 12.9 | 36.8 | 13 - 16 |
| Q2: What are the signs and symptoms of colon cancer? | 12.2 | 41.3 | 13 - 16 |
| Q3: What are the signs and symptoms of liver cancer? | 11.8 | 41.2 | 13 - 16 |
| Q4: What are the different stages and treatments for pancreatic cancer? | 13.4 | 28.6 | College graduate |
| Q5: What are the different stages and treatments for colon cancer? | 15.1 | 16.4 | College graduate |
| Q6: What are the different stages and treatments for liver cancer? | 14.5 | 17.8 | College graduate |
| Q7: Who is appropriate for surgery for pancreatic cancer? | 18.1 | 20.6 | College graduate |
| Q8: Who is appropriate for surgery for colon cancer? | 18.0 | 17.4 | College graduate |
| Q9: Who is appropriate for surgery for liver cancer? | 16.6 | 21.5 | College graduate |
| Q10: How long is the recovery from pancreatic cancer surgery? | 16.1 | 20.0 | College graduate |
| Q11: How long is the recovery from colon cancer surgery? | 16.2 | 11.7 | College graduate |
| Q12: How long is the recovery from liver cancer surgery? | 14.5 | 17.4 | College graduate |
| Q13. What are the risks of surgery to remove my pancreatic cancer? | 14.8 | 20.7 | College graduate |
| Q14. What are the risks of surgery to remove my colon cancer? | 14.2 | 22.8 | College graduate |
| Q15. What are the risks of surgery to remove my liver cancer? | 14.9 | 15.6 | College graduate |
| Q16. Will I need an ostomy after surgery to remove my colon cancer? | 14.2 | 32.9 | 13 - 16 |
| Q17: How long will I be in the hospital after surgery for cancer in my belly? | 14.7 | 36.2 | 13 - 16 |
| Q18: How long after surgery for cancer in my belly can I exercise? | 15.4 | 26.5 | College graduate |
| Q19: How long will it take to recover from surgery for cancer in my belly? | 14.1 | 31.8 | 13 - 16 |
| Q20: Should I stay close to the hospital in a hotel or Airbnb after I'm discharged from surgery for cancer in my belly? | 17.0 | 32.7 | 13 - 16 |
| Q21: How long after surgery for cancer in my belly can I do chores around the house? | 15.3 | 33.9 | 13 - 16 |
| Q22: I just had surgery for cancer in my belly, and my incision is painful. What do I do? | 12.1 | 43.6 | 13 - 16 |

| ID and question | FKRGL score (US-grade reading level) | FRE score | Estimated FRE US-grade level [34] |
|---|---|---|---|
| Q23: I just had surgery for cancer in my belly, and I am still in some pain. Is there anything else I can take for the pain? | 12.7 | 36.6 | 13 - 16 |
| Q24: I just had surgery for cancer in my belly, and my incision is starting to hurt more and looks slightly open. What do I do? | 10.8 | 48.0 | 13 - 16 |
| Q25: I just had surgery for cancer in my belly, and the incision is warm to the touch and draining a yellowish fluid. What do I do? | 13.6 | 40.0 | 13 - 16 |
| Q26: I just had surgery for cancer in my belly. It hurts when I breathe, and I have a new cough. What do I do? | 16.2 | 27.4 | College graduate |
| Q27: I just had surgery for cancer in my belly. It now burns when I pee. What do I do? | 15.0 | 30.1 | 13 - 16 |
| Q28: I am about to have surgery for cancer in my belly. How can I prevent an infection after? | 12.0 | 36.1 | 13 - 16 |

[a]FKRGL: Flesch-Kincaid Reading Grade Level.

[b]FRE: Flesch Reading Ease.

Four questions were selected for post hoc analysis shown in Table 4 (Q1, Q7, Q11, and Q15). These questions had an original median FKRGL score of 15.6 (IQR 14.4-16.7; range 12.9 - 18.1) and FRE score of 18.1 (IQR 14.6-24.7; range: 11.7 - 36.8). When GPT-4o was queried to respond to select questions to the level of a 5th-grade reader, the median FKRGL score decreased to 7.1 (IQR 6.1-8.3; range: 5.9 - 9.0) and FRE increased to 73.8 (IQR 66.6-79.3; range: 60.1 - 80.9). Two independent graders (CVL and DAS) found responses to result in a mean content similarity of 3.88 (SD 0.25) in comparison to the original response. Responses to questions that were rephrased by GPT, with prompting to query at a 5th-grade reading level, resulted in a median FKRGL score of 14.5 (IQR 13.2-15.4; range: 11.6 to 15.8), a median FRE score of 32.0 (IQR 27.0-37.7; range: 21.4 - 45.0), and a mean content similarity of 4.63 (SD 0.25) to original responses. The raters had identical scores for 50% (4/8) of responses, with the other 4 responses differing by 1 on the 5-point Likert scale.

**Table .** FKRGL[a] and FRE[b] scores for select questions (V1), questions prompted to respond at the 5th-grade level (V2), and questions rephrased by GPT to be asked at the 5th-grade level (V3).

| Question and versions | FKRGL score | FRE score | Content similarity, mean (SD) |
|---|---|---|---|
| Q1 | | | |
| V1: What are the signs and symptoms of pancreatic cancer? | 12.9 | 36.8 | Reference |
| V2: Answer at a 5th-grade level: What are the signs and symptoms of pancreatic cancer? | 6.1 | 80.9 | 4.0 (0) |
| V3: What are the warning signs of pancreatic cancer and how might someone feel if they have it? | 11.6 | 45.0 | 4.5 (0.71) |
| Q7 | | | |
| V1: Who is appropriate for surgery for pancreatic cancer? | 18.1 | 20.6 | Reference |
| V2: Answer at a 5th-grade level: Who is appropriate for surgery for pancreatic cancer? | 9.0 | 60.1 | 4.0 (0) |
| V3: Who can have surgery to treat pancreatic cancer? | 15.3 | 28.8 | 5.0 (0) |
| Q11 | | | |
| V1: How long is the recovery from colon cancer surgery? | 16.2 | 11.7 | Reference |
| V2: Answer at a 5th-grade level: How long is the recovery from colon cancer surgery? | 5.9 | 78.8 | 3.5 (0.71) |
| V3: How long does it take to feel better after colon cancer surgery? | 15.8 | 21.4 | 4.5 (0.71) |
| Q15 | | | |
| V1: What are the risks of surgery to remove my liver cancer? | 14.9 | 15.6 | Reference |
| V2: Answer at a 5th-grade level: What are the risks of surgery to remove my liver cancer? | 8.1 | 68.8 | 4.0 (0) |
| V3: What could go wrong if I have surgery to take out my liver cancer? | 13.7 | 35.2 | 4.5 (0.71) |

[a]FKRGL: Flesch-Kincaid Reading Grade Level.

[b]FRE: Flesch Reading Ease.

## Qualitative Feedback

Numerous inaccuracies within GPT-generated responses were detected by a surgeon-expert concerning general disease information and postoperative recovery. The following feedback has been modified for clarity but maintains the original intent. In Q2 (signs and symptoms of colon cancer), rectal bleeding was mistakenly described as a systemic symptom, while it is a local symptom that may lead to secondary systemic symptoms, including fatigue due to anemia. For Q10 (pancreatic cancer surgery recovery), "light activities," which are often defined as walking or activities of daily living in the surgical setting, were resumed while a patient was admitted, instead of the written 6 to 12 weeks following discharge. Likewise, for Q11 (colon cancer surgery recovery), certain "light activities" could be resumed sooner. For Q12 (liver cancer surgery recovery), the mention of major hepatectomy as treatment was notably absent.

Regarding quality, numerous content gaps were noted. For Q3 (signs and symptoms of liver cancer), the response described chronic liver disease symptoms; these are common for patients with primary liver cancers but less frequent in the setting of secondary liver cancers (ie, colorectal cancer with liver metastases). Regarding Q4 (stages and treatments of pancreatic cancer), genetic testing should be included when discussing targeted therapies. For Q7 (pancreatic cancer surgery eligibility), discussion of the biology of resectability, which is accounted for by tumor markers such as Ca 19‐9, was notably absent. Regarding Q20 (staying near the hospital following discharge), while listed, it is not emphasized that staying nearby is unnecessary unless the patient lives far away. Furthermore, the

question could be enhanced by including discussion of local housing options with case management or a social worker. For Q25 (postoperative infection), concern for dehiscence is not explicitly stated, and the volume of drainage should be addressed sooner, as high volume may indicate dehiscence.

### Provider Recommendations for GPT as a Patient Resource

When assessed on their comfort level with patients using publicly available AI for health information, 57.1% (4/7) of providers reported being "comfortable," 14.3% (1/7) reported being "neither comfortable nor uncomfortable," 14.3% (1/7) reported being "uncomfortable," and 14.3% (1/7) reported being "very uncomfortable." Regarding provider discomfort, when asked for questions or concerns pertaining to the study, 1 respondent reported "The answers should be designed for a lower health literacy level." Another physician expressed concern over direct patient use of ChatGPT, primarily citing lack of supervision and noting "health is not something you want to leave up to a robot. There will always be intricacies that cannot be understood by AI."

## Discussion

### Principal Findings

This study is among the first to evaluate ChatGPT-4o as a patient information resource for individuals preparing for or recovering from surgery for abdominal malignancies [20,35]. As patient self-use of LLMs for medical information is increasing [14], it is essential to assess the content quality, safety, and comprehensibility of GPT-generated responses. Through gaining a deeper understanding of the strengths and weaknesses present within LLMs, providers may help patients be aware of such options and help them navigate the use of these sources. The current study's results indicate that ChatGPT-4o may serve as a useful patient information resource, with most responses rated from "good" to "very good to excellent" in quality, accuracy, and tangibility. Notably, the lowest rated responses received a median score of 3.0, corresponding to a "good" rating, whereas the highest rated responses received a score of 4.50, corresponding to a rating between "very good" and "excellent." However, there is still room for improvement in generated responses prior to the endorsement of ChatGPT as a "gold-standard" patient resource. While most providers were "comfortable" having patients use publicly available AI for health information, 42.9% (3/7) of providers did not report feeling "comfortable" having patients use publicly available AI for health information, with 2 reported being "uncomfortable" or "very uncomfortable." Physicians cited concerns regarding patient use of ChatGPT, noting poor response comprehensibility and lack of supervision, factors likely contributing to their lack of comfort in patient use of ChatGPT. Moreover, this study raises concerns about the comprehensibility of the generated responses, as elevated FKGRL scores indicate that many require a postsecondary reading level for adequate understanding. Physicians should be aware that, in the context of patient use of LLMs for medical information, patients would benefit from instructions for use and monitoring for potential ChatGPT-derived misconceptions.

Information is scarce regarding the safety and accuracy of ChatGPT-generated responses in the perioperative setting for abdominal malignancies. Given the complex biological mechanisms and therapeutic management of gastrointestinal malignancies, it is critical to evaluate the quality of ChatGPT-generated content. The presented data suggest that, although ChatGPT responses averaged as "good" or "very good," scores were highly question- and domain-dependent. Given its high overall ratings, ChatGPT may serve as an advantageous tool for patients to develop a baseline knowledge of their disease prior to clinical encounters. However, numerous inaccuracies and content gaps were identified within responses. This is congruent with past work assessing ChatGPT's use for thoracic surgery, where most responses likewise ranged from "good" to "very good," minor inaccuracies were identified in each answer, and certain domains performed better than others [32]. Regarding abdominal malignancies, questions concerning staging and treatment received the lowest mean accuracy and quality scores. As such, providers should be encouraged to assess potential disease misconceptions that patients using ChatGPT may have and ensure they distribute comprehensive general disease information. Interestingly, ChatGPT-4o excelled in answering questions pertaining to adverse outcomes following surgery. As such, ChatGPT may help guide patients seeking proper management for postoperative complications.

The present study suggests that patient information regarding abdominal malignancies presented by ChatGPT-4o may produce material that is poorly comprehensible for many of the intended population due to requirements of high health literacy and education level. One surgeon expressed that responses should be written for a lower health literacy level. This is consistent with the findings of a high grade level requirement for adequate comprehensibility (FKRGL score), averaging a grade level of 14.5. An FKRGL of 14.5 represents a reading level requiring some level of college education. Current recommendations suggest that patient resources should be tailored to a 5th-grade reading level for accessibility [36,37].

Readability as a limitation of ChatGPT has been previously reported in the literature [15,22]. Past work regarding cervical spine surgery likewise noted high FKRGL scores to limit ChatGPT-3.5's use as a patient resource. After prompting ChatGPT to provide answers at a 6th-grade reading level, answers decreased from a grade level of 13.5 to 11.2, though remaining persistently elevated. Notably, the present study used similar techniques that successfully produced responses at a lower reading level with ChatGPT-4o. For select questions, the median FKRGL score prior to rephrasing or prompting was 15.6 (IQR 14.4-16.7). Remarkable improvement was noted upon prompting GPT to respond to the level of a 5th-grade reader, decreasing the median FKRGL score, or US grade level, to 7.1 (IQR 6.1-8.3). This work suggests that improvements within ChatGPT-4o may allow for more comprehensible responses, given appropriate prompting. Notably, there was a less meaningful drop in median FKRGL score, from 15.6 (IQR 14.4-16.7) to 14.5 (IQR 13.2-15.4), when ChatGPT was used to *rephrase* questions to be asked at the reading level of a 5th grader. Prior to modification, the median FRE score for select questions was 18.1 (IQR 14.6-24.7). Consistent with FKRGL

trends, prompting questions resulted in a more substantial increase in FRE score (median FRE 73.8, IQR 66.6-79.3), indicating markedly improved readability, compared with rephrasing (median FRE 32, IQR 27.0-37.7). This suggests that explicitly requesting ChatGPT to produce responses at a lower level may be more effective in improving readability than adjusting question phrasing. Although contents similar to original questions were better for the latter group than the former (4.63/5 versus 3.88/5), most key concepts were retained within both groups. As such, prompting ChatGPT to answer at a lower grade level may improve readability without significantly sacrificing content. Therefore, providers should be encouraged to assess patient use of LLMs for medical questions and provide patients with a menu for how to prompt ChatGPT to answer at an appropriate grade level if relevant.

While comprehensibility without prompting educational level can be a limitation for the intended patient population, ChatGPT may serve as a useful tool for providers and trainees. Past work in public health has found AI chatbots to be a useful educational tool for medical students in answering complex medical questions [38]. Within the present study, questions 1 to 9 pertain to “signs and symptoms,” “stages and treatment,” and “surgery eligibility”; these questions may be asked by clinicians or learners. While patient readability was limited by a high mean grade level required, the ratings typically ranged from “good” to “very good” in quality, accuracy, and tangibility. This suggests that ChatGPT-4o can serve as a useful resource for physicians and medical trainees, given a higher health literacy than the general population. To further evaluate the use and comprehensibility of ChatGPT as a patient resource, future investigations should involve patient perspectives.

This study has several limitations. First, the small sample size (n=7) of surgical oncologists grading the responses substantially limits statistical power and the reliability of the findings. The single-institutional nature of the study further limits generalizability, as physician responses may reflect regional practice patterns and institutional biases. Future validation should evaluate larger, multi-institutional cohorts to confirm reproducibility and evaluate external validity. Second, the survey incorporated subjective assessments which may limit reproducibility, as concepts graded, such as “quality,” “accuracy,” and “tangibility,” are abstract. To enhance reproducibility, standardized definitions of these domains were included on each page of the survey. Third, questions may not be well representative of patient language. Although questions were obtained from hospital websites and piloted with residents to improve alignment with patient phrasing, they may not encompass the full spectrum of patient inquiries nor the variability of patients’ health literacy. As only 28 questions were assessed across 3 malignancies, the nature of the questions is limited in scope and may not represent all questions patients may ask pertaining to their diagnosed malignancy. Moreover, questions are broad, pertaining to “colon,” “pancreas,” “liver,” or “belly” cancers, without specifying types and stages.

## Conclusions

This preliminary study indicates that, while publicly accessible ChatGPT may serve as a useful patient resource, its use as an unsupervised source of information for patients with abdominal malignancies has distinct limitations. Providers should be aware that many of their patients are accessing ChatGPT and recognize that developing an understanding of its strengths and limitations can help them guide their patients to enable its best use. Inaccuracies, gaps in information, and poor readability were identified in ChatGPT-generated content, suggesting patients may benefit from physician guidance. Providers should be prepared to properly support their patients reporting ChatGPT use by counseling techniques such as prompting questions to tailor responses to their educational level. The data herein indicate that this is critical for the interpretation of the information by patients, as without this guidance, the answers are directed to an educational level of college or above.

## Acknowledgments

During the preparation of this work, the authors used ChatGPT-4o as specified above to collect the information required for data analysis (artificial intelligence–generated responses). No generative artificial intelligence was used in the writing process.

## Funding

The authors declared no financial support was received for this work.

## Data Availability

To preserve respondent confidentiality, the data set used in this study is not publicly available. A limited version of the dataset used in this study can be obtained from the first author upon request.

## Authors' Contributions

Conceptualization: CVL, DAS, ANM, CLC, KNS, MEL, DPN, RS
Data curation: CVL, DAS, ANM, CLC, KNS, MEL, DPN, RS
Formal analysis: CVL, DAS, RS
Methodology: CVL, DAS, RS
Supervision: ANM, CLC, KNS, MEL, DPN, RS
Writing – original draft: CVL, RS

Writing – review & editing: CVL, DAS, ANM, CLC, KNS, MEL, DPN, RS

## Conflicts of Interest

RS serves on the Board of Directors of DNAnexus, Heartland Whole Health Institute, ZealCare, Inc, where he is also the cofounder; Board of Trustees of American Medical Program, Tel Aviv University, and Scientific Advisory Board of OrthoBioTherapeutics Inc.

Multimedia Appendix 1
Distributed survey with GPT-generated responses.
[DOCX File, 31 KB - periop_v9i1e81374_app1.docx ]

## References


1. ChatGPT. URL: https://chatgpt.com [accessed 2025-03-09]
2. Guinness H. How does ChatGPT work? Zapier. URL: https://zapier.com/blog/how-does-chatgpt-work/ [accessed 2025-04-15]
3. Bubeck S, Chandrasekaran V, Eldan R, Gehrke J, Horvitz E, Kamar E, et al. Sparks of artificial general intelligence: early experiments with GPT-4. arXiv. Preprint posted online on Mar 22, 2023. [doi: 10.48550/arXiv.2303.12712]
4. Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLOS Digit Health 2023 Feb;2(2):e0000198. [doi: 10.1371/journal.pdig.0000198] [Medline: 36812645]
5. Gilson A, Safranek CW, Huang T, et al. How does ChatGPT perform on the United States Medical Licensing Examination (USMLE)? The implications of large language models for medical education and knowledge assessment. JMIR Med Educ 2023 Feb 8;9:e45312. [doi: 10.2196/45312] [Medline: 36753318]
6. Gupta R, Herzog I, Park JB, et al. Performance of ChatGPT on the plastic surgery inservice training examination. Aesthet Surg J 2023 Nov 16;43(12):NP1078-NP1082. [doi: 10.1093/asj/sjad128] [Medline: 37128784]
7. Hoch CC, Wollenberg B, Lüers JC, et al. ChatGPT's quiz skills in different otolaryngology subspecialties: an analysis of 2576 single-choice and multiple-choice board certification preparation questions. Eur Arch Otorhinolaryngol 2023 Sep;280(9):4271-4278. [doi: 10.1007/s00405-023-08051-4] [Medline: 37285018]
8. Mihalache A, Huang RS, Popovic MM, Muni RH. Performance of an upgraded artificial intelligence chatbot for ophthalmic knowledge assessment. JAMA Ophthalmol 2023 Aug 1;141(8):798-800. [doi: 10.1001/jamaophthalmol.2023.2754] [Medline: 37440220]
9. Cabral S, Restrepo D, Kanjee Z, et al. Clinical reasoning of a generative artificial intelligence model compared with physicians. JAMA Intern Med 2024 May 1;184(5):581-583. [doi: 10.1001/jamainternmed.2024.0295] [Medline: 38557971]
10. Tan S, Xin X, Wu D. ChatGPT in medicine: prospects and challenges: a review article. Int J Surg 2024 Jun 1;110(6):3701-3706. [doi: 10.1097/JS9.0000000000001312] [Medline: 38502861]
11. Liu HY, Alessandri Bonetti M, De Lorenzi F, Gimbel ML, Nguyen VT, Egro FM. Consulting the digital doctor: Google versus ChatGPT as sources of information on breast implant-associated anaplastic large cell lymphoma and breast implant illness. Aesthetic Plast Surg 2024 Feb;48(4):590-607. [doi: 10.1007/s00266-023-03713-4] [Medline: 37903939]
12. Bergmo TS, Sandsdalen V, Manskow US, Småbrekke L, Waaseth M. Internet use for obtaining medicine information: cross-sectional survey. JMIR Form Res 2023 Feb 2;7:e40466. [doi: 10.2196/40466] [Medline: 36729577]
13. Ramli R, Jambor MA, Kong CY. Dr Google - assessing the reliability and readability of information on general surgical procedures found via search engines. ANZ J Surg 2023 Mar;93(3):590-596. [doi: 10.1111/ans.18289] [Medline: 36716246]
14. Ayre J, Cvejic E, McCaffery KJ. Use of ChatGPT to obtain health information in Australia, 2024: insights from a nationally representative survey. Med J Aust 2025 Mar 3;222(4):210-212. [doi: 10.5694/mja2.52598] [Medline: 39901778]
15. Shen SA, Perez-Heydrich CA, Xie DX, Nellis JC. ChatGPT vs. web search for patient questions: what does ChatGPT do better? Eur Arch Otorhinolaryngol 2024 Jun;281(6):3219-3225. [doi: 10.1007/s00405-024-08524-0] [Medline: 38416195]
16. Artioli E, Veronesi F, Mazzotti A, et al. Assessing ChatGPT responses to common patient questions regarding total ankle arthroplasty. J Exp Orthop 2024;12(1):e70138. [doi: 10.1002/jeo2.70138] [Medline: 39741912]
17. Gajjar AA, Kumar RP, Paliwoda ED, et al. Usefulness and accuracy of artificial intelligence chatbot responses to patient questions for neurosurgical procedures. Neurosurgery 2024 Feb 14;95(1). [doi: 10.1227/neu.0000000000002856] [Medline: 38353558]
18. Samaan JS, Yeo YH, Rajeev N, et al. Assessing the accuracy of responses by the language model ChatGPT to questions regarding bariatric surgery. Obes Surg 2023 Jun;33(6):1790-1796. [doi: 10.1007/s11695-023-06603-5] [Medline: 37106269]
19. Yeo YH, Samaan JS, Ng WH, et al. Assessing the performance of ChatGPT in answering questions regarding cirrhosis and hepatocellular carcinoma. Clin Mol Hepatol 2023 Jul;29(3):721-732. [doi: 10.3350/cmh.2023.0089] [Medline: 36946005]
20. Rydzewski NR, Dinakaran D, Zhao SG, et al. Comparative evaluation of LLMs in clinical oncology. NEJM AI 2024 May;1(5). [doi: 10.1056/aioa2300151] [Medline: 39131700]
21. Lee TC, Staller K, Botoman V, Pathipati MP, Varma S, Kuo B. ChatGPT answers common patient questions about colonoscopy. Gastroenterology 2023 Aug;165(2):509-511.e7. [doi: 10.1053/j.gastro.2023.04.033] [Medline: 37150470]

22. Subramanian T, Araghi K, Amen TB, et al. Chat generative pretraining transformer answers patient-focused questions in cervical spine surgery. Clin Spine Surg 2024 Jul 1;37(6):E278-E281. [doi: 10.1097/BSD.0000000000001600] [Medline: 38531823]
23. Al-Dujaili Z, Omari S, Pillai J, Al Faraj A. Assessing the accuracy and consistency of ChatGPT in clinical pharmacy management: a preliminary analysis with clinical pharmacy experts worldwide. Res Social Adm Pharm 2023 Dec;19(12):1590-1594. [doi: 10.1016/j.sapharm.2023.08.012] [Medline: 37696742]
24. Arnold M, Abnet CC, Neale RE, et al. Global burden of 5 major types of gastrointestinal cancer. Gastroenterology 2020 Jul;159(1):335-349.e15. [doi: 10.1053/j.gastro.2020.02.068] [Medline: 32247694]
25. Pancreatic cancer questions to ask the healthcare team. Pancreatic Cancer Action Network. 2025. URL: https://pancan.org/facing-pancreatic-cancer/diagnosis/choosing-your-healthcare-team/questions-to-ask-your-healthcare-team [accessed 2025-01-15]
26. Colon cancer FAQs. Moffitt Cancer Center. 2025. URL: https://www.moffitt.org/cancers/colon-cancer/faqs [accessed 2025-01-15]
27. Frequently asked questions. Mount Sinai Tisch Cancer Center. 2025. URL: https://www.mountsinai.org/care/cancer/services/colon/faqs [accessed 2025-01-15]
28. Liver cancer – frequently asked questions. Pelican Cancer Foundation. 2025. URL: https://www.pelicancancer.org/patient-resources/liver-cancer/frequently-asked-questions [accessed 2025-01-15]
29. Questions to ask about pancreatic cancer. American Cancer Society. 2024. URL: https://www.cancer.org/cancer/types/pancreatic-cancer/detection-diagnosis-staging/talking-with-doctor.html [accessed 2025-01-15]
30. Questions to ask about liver cancer. American Cancer Society. 2025. URL: https://www.cancer.org/cancer/types/liver-cancer/detection-diagnosis-staging/talking-with-doctor.html [accessed 2025-01-15]
31. Frequently asked questions. Hirshberg Foundation for Pancreatic Cancer Research. 2025. URL: https://pancreatic.org/pancreatic-cancer/faqs [accessed 2025-01-15]
32. Ferrari-Light D, Merritt RE, D'Souza D, et al. Evaluating ChatGPT as a patient resource for frequently asked questions about lung cancer surgery-a pilot study. J Thorac Cardiovasc Surg 2025 Apr;169(4):1174-1180. [doi: 10.1016/j.jtcvs.2024.09.030] [Medline: 39326732]
33. ChatGPT — release notes. OpenAI. 2026. URL: https://help.openai.com/en/articles/6825453-chatgpt-release-notes?utm_source=chatgpt.com [accessed 2026-03-01]
34. Jindal P, MacDermid JC. Assessing reading levels of health information: uses and limitations of flesch formula. Educ Health (Abingdon) 2017;30(1):84-88. [doi: 10.4103/1357-6283.210517] [Medline: 28707643]
35. Munir MM, Endo Y, Ejaz A, Dillhoff M, Cloyd JM, Pawlik TM. Online artificial intelligence platforms and their applicability to gastrointestinal surgical operations. J Gastrointest Surg 2024 Jan;28(1):64-69. [doi: 10.1016/j.gassur.2023.11.019] [Medline: 38353076]
36. AHRQ health literacy universal precautions toolkit. : Agency for Healthcare Research and Quality (AHRQ); 2024 URL: https://www.ahrq.gov/sites/default/files/wysiwyg/health-literacy/3rd-edition-toolkit/health-literacy-toolkit-third-edition.pdf [accessed 2025-04-20]
37. Stossel LM, Segar N, Gliatto P, Fallar R, Karani R. Readability of patient education materials available at the point of care. J Gen Intern Med 2012 Sep;27(9):1165-1170. [doi: 10.1007/s11606-012-2046-0] [Medline: 22528620]
38. Baglivo F, De Angelis L, Casigliani V, Arzilli G, Privitera GP, Rizzo C. Exploring the possible use of AI chatbots in public health education: feasibility study. JMIR Med Educ 2023 Nov 1;9:e51421. [doi: 10.2196/51421] [Medline: 37910155]


## Abbreviations

**AI:** artificial intelligence
**FKRGL:** Flesch-Kincaid Reading Grade Level
**FRE:** Flesch Reading Ease
**LLM:** large language model
**UTI:** urinary tract infection

*Edited by J Shiffermiller; submitted 27.Jul.2025; peer-reviewed by C Richlitzki, A Roland, S Matsuda; revised version received 08.Mar.2026; accepted 12.Mar.2026; published 04.May.2026.*

*Please cite as:*
*Lindsay CV, Shenoy DA, Martin AN, Clipper CL, Shah KN, Lidsky ME, Nussbaum DP, Snyderman R*
*Physician Perspectives on ChatGPT-4o as a Patient Resource for Abdominal Cancer Surgeries: Cross-Sectional Survey*
*JMIR Perioper Med 2026;9:e81374*
*URL: https://periop.jmir.org/2026/1/e81374*
*doi:10.2196/81374*

# AI-Generated Avatar Videos for Postoperative Patient Education Among Health Care Workers: Pilot Randomized Controlled Trial

Syed Ali Haider[1], MBBS; Srinivasagam Prabha[1], PhD; Cesar Abraham Gomez-Cabello[1], MD; Ariana Genovese[1], BS; Bernardo G Collaco[1], MD; Nadia Wood[2], MS; James London[3], MHA; Sanjay Bagaria[4], MD; Mark A Lifson[5], PhD; Cui Tao[6], PhD; Antonio Jorge Forte[1,5,6], MD, PhD

[1]Department of Surgery, Division of Plastic Surgery, Mayo Clinic in Florida, 4500 San Pablo Road, Jacksonville, FL, United States

[2]Department of Radiology, AI IT, Mayo Clinic, Rochester, MN, United States

[3]Department of Surgery, Mayo Clinic in Florida, Jacksonville, FL, United States

[4]Division of Surgical Oncology, Mayo Clinic, Jacksonville, FL, United States

[5]Center for Digital Health, Mayo Clinic, Rochester, MN, United States

[6]Department of Artificial Intelligence and Informatics, Mayo Clinic in Florida, Jacksonville, FL, United States

**Corresponding Author:**
Antonio Jorge Forte, MD, PhD
Department of Surgery, Division of Plastic Surgery, Mayo Clinic in Florida, 4500 San Pablo Road, Jacksonville, FL, United States

## Abstract

**Background:** Effective postoperative communication is vital for patient recovery, yet traditional text-based discharge instructions often lead to poor comprehension and adherence, particularly among patients with limited health literacy. Although educational videos improve understanding and retention, their widespread use has been hampered by high production costs. Generative artificial intelligence (AI) offers a scalable solution for creating engaging video content.

**Objective:** The primary objective of this pilot study was to assess the feasibility of creating and deploying AI-generated, avatar-led videos for postoperative instruction delivery. Secondary objectives included comparing knowledge retention, engagement, perceived clarity, and user experience between AI-generated video and traditional text-based handout formats among health care workers.

**Methods:** In this randomized pilot study, 38 health care worker volunteers were recruited as a convenience sample to pilot-test the intervention before patient implementation. Participants were assigned to either a text handout group (n=19, 50%) or an AI-generated video group (n=19, 50%). Both groups received information on 10 common postoperative topics. The primary outcome was objective knowledge, assessed via a 10-item quiz. Secondary outcomes, measured through surveys with 5-point Likert scales, included engagement time, subjective engagement, perceived clarity, usefulness, confidence in understanding, and information retention. Qualitative feedback was also collected.

**Results:** Objective knowledge quiz scores did not differ significantly between groups (mean 8.89, SD 1.20 for the AI-generated video group vs mean 8.21, SD 1.78 for the text handout group; $P$=.17; Cohen $d$=0.45). Participants in the AI-generated video group demonstrated significantly higher engagement time (mean 15.11, SD 7.78 minutes vs mean 8.84, SD 4.03 minutes; $P$=.004; Cohen $d$=1.04). They also rated instructions as significantly clearer (mean 4.63, SD 0.50 vs mean 4.00, SD 0.82; $P$=.007; Cohen $d$=0.93), more engaging (mean 4.05, SD 0.78 vs mean 3.32, SD 1.00; $P$=.02; Cohen $d$=0.81), and more effective for retention (mean 4.42, SD 0.84 vs mean 3.37, SD 0.68; $P$<.001; Cohen $d$=1.38). Qualitative feedback highlighted the engaging nature of AI-generated videos but noted areas for avatar refinement.

**Conclusions:** In this pilot study with health care workers, AI-generated avatar videos did not improve objective knowledge scores but significantly enhanced engagement, perceived retention and perceived clarity (Cohen $d$=0.81 - 1.38). Future studies in actual patient populations with diverse health literacy levels are needed to determine whether these engagement advantages translate into improved knowledge outcomes.

***(JMIR Perioper Med 2026;9:e89277)*** doi:10.2196/89277

**KEYWORDS**

artificial intelligence; AI; text to video; generative artificial intelligence; generative AI; postoperative instructions; patient education

## Introduction

Effective postoperative communication at hospital discharge is critical, yet many patients leave with written instructions they cannot fully understand or remember. Approximately 90 million adults in the United States have limited health literacy [1]. Patients with lower educational attainment may struggle to recall instructions, leading to medication errors, missed follow-ups, and adverse outcomes [2-5]. Poor adherence to discharge instructions contributes to preventable readmissions, patient dissatisfaction, and increased health care costs [6-8]. Traditional instruction methods, such as written materials and brief verbal explanations, rarely deliver information in a memorable or actionable way [9-11].

Educational videos offer a compelling solution [12,13]. They simplify complex information, support diverse learning styles, and improve comprehension and retention—especially when paired with written materials [12]. Videos can demonstrate procedures visually, feature presenters who create a sense of social presence, accommodate varying literacy levels through combined auditory and visual information, and offer the potential for multilingual messaging. However, high production costs, long turnaround times, and equipment requirements have limited their widespread use.

Generative artificial intelligence (AI) enables the rapid, automated creation of video content [13]. Platforms such as OpenAI's Sora and Google DeepMind's Veo can generate high-quality, coherent video content directly from simple text prompts [14], with applications proposed across medical domains, including surgical training, public health education, and patient-specific educational content [15-17]. A particularly compelling feature is AI-generated avatars—photorealistic digital humans who deliver content in a natural manner.

According to Mayer's cognitive theory of multimedia learning [18], combining verbal and visual information enhances understanding by engaging dual channels in the brain. Visible instructors create social presence, increasing motivation and attention [19]. A systematic review found that videos featuring visible instructors consistently improved learning outcomes [20]. Early evidence in health care supports this: a virtual nurse avatar for discharge instructions was well received, with 85% of patients liking the character and 86.4% of nurses believing it would aid education [21]. AI-generated avatars have been explored for HIV education, heart disease knowledge dissemination, and ileostomy education [22-24], and more recently, interactive physician avatars for postoperative patient education have demonstrated high usability and trust among surgical patients [25].

By assessing both objective understanding and subjective experience, this study aims to examine the feasibility and user experience of AI-generated video as a medium for patient education at hospital discharge. If effective, this approach could help reduce disparities in postoperative care by providing patients with clearer, more accessible discharge instructions.

## Methods

### Study Design and Participants

The primary objective was to assess the feasibility of creating and deploying AI-generated avatar videos for postoperative instruction. The primary outcome was the difference in knowledge quiz scores between groups. Secondary outcomes included engagement time, subjective engagement, perceived clarity, usefulness, confidence in understanding, information retention, and preference for format. Qualitative feedback was also collected through open-ended questions.

We conducted a pilot study to compare traditional text-based instructions with AI-generated video instructions for postoperative care education. A total of 38 adult health care worker volunteers were recruited from a hospital community (faculty, staff, researchers, and students). This convenience sample of health care workers was chosen (1) to pilot-test the intervention in a controlled setting before patient implementation and (2) to establish feasibility and refine study procedures prior to recruiting actual patients.

Participants were recruited through email announcements sent to hospital faculty, staff, researchers, and students. Interested volunteers contacted the research team via email and were provided with study information. Inclusion criteria were (1) adult health care workers (aged ≥18 years), (2) proficiency in English, and (3) access to a computer or mobile device with internet connectivity. Exclusion criteria included (1) prior involvement in postoperative instruction material development and (2) visual or hearing impairments that would prevent video viewing or text reading. All instruments were administered online using Google Forms in a single session. Participants accessed the form through a unique link and completed all sections sequentially.

### Ethical Considerations

This study was reviewed and approved by the Mayo Clinic Institutional Review Board (application 25-002248). The study was conducted in accordance with the ethical standards of the responsible institutional committee on human experimentation and with the World Medical Association's Declaration of Helsinki.

All participants were adult health care worker volunteers (faculty, staff, researchers, and students) recruited through institutional email announcements. Written informed consent was obtained electronically through the Google Forms platform before randomization and data collection began. Participants were informed of the study's purpose and voluntary nature and their right to withdraw at any time without consequences.

Participant privacy and confidentiality were maintained throughout the study. All survey responses were collected anonymously via Google Forms; no personally identifiable information (eg, name, employee ID, email address, or date of birth) was linked to individual survey responses. Deidentified study data were stored on password-protected, institutionally managed servers accessible only to authorized members of the research team. No images, audio, or video of participants were

collected. Only aggregate, deidentified data are reported in this manuscript.

Participants received no financial compensation, gift cards, or other incentives for their participation in this study.

## Educational Materials

### *Text-Based Instructions*

We compiled a set of 10 standard postoperative instruction handouts covering common topics relevant to general surgery recovery. These were developed based on established postoperative care guidelines and publicly available patient education resources from major medical institutions [26] and included the following 10 frequently asked topics: pain, pain medication, alarm signs after surgery, surgical drain management, diet after surgery, postoperative nausea, follow-up appointment instructions, recovery instructions, scars after surgery, and suture management [27-36]. Each handout was approximately 100 to 200 words in length and written at an eighth- to ninth-grade reading level [37]. These served as the standard care educational materials for the control group.

### *AI-Generated Video Instructions*

Using HeyGen (HeyGen Inc) [38], we created video versions covering the same topics. Scripts were based on the text handout content and edited for a conversational tone. Average video duration was 59.2 (range 45 - 84) seconds. We used realistic, diverse avatar presenters (3 male and 3 female avatars) with AI-generated voices. The pain and pain medication topics were merged, yielding 9 videos covering 10 topics. Videos featured avatars speaking directly to viewers, with bullet points and subtitles displayed (Figure 1 [25]).

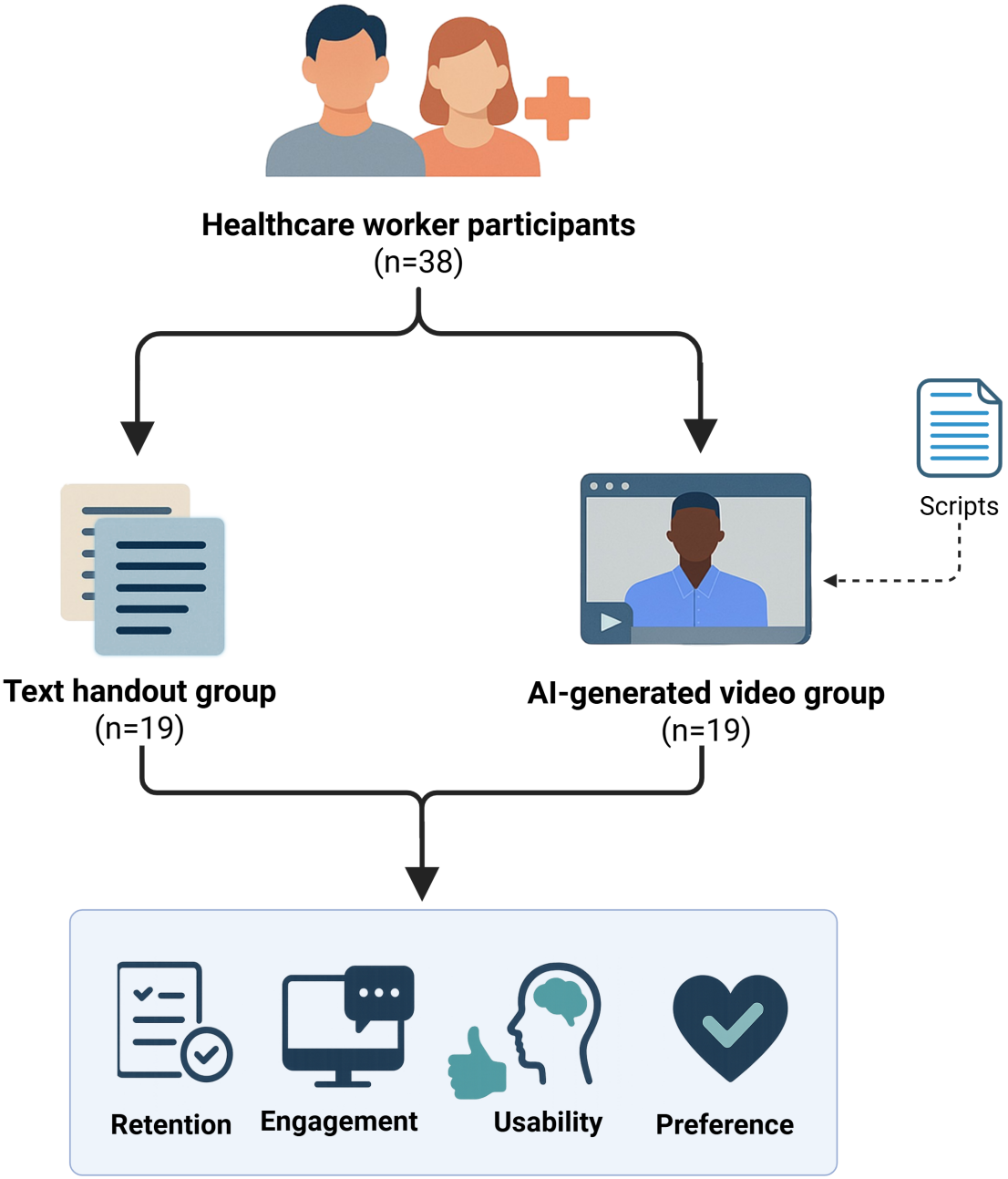
**Figure 1.** Examples of text-to-video avatars demonstrating the generative artificial intelligence's ability to generate realistic and diverse human-like representations. Images were generated using HeyGen (HeyGen Inc). Collage created with BioRender.com [39].

From a practical standpoint, the HeyGen platform enabled rapid video production without specialized equipment, studios, or video editing expertise. Each video was generated within minutes of script input, and the total time to produce all 9 videos was approximately 2 to 3 hours, including script preparation and quality review.

## Randomization

Volunteers were randomly assigned in a 1:1 ratio to either the text handout group (control) or the AI-generated video group (intervention) using simple randomization (drawing lots). Randomization was performed after informed consent was obtained and before participants accessed the educational materials (Figure 2).

**Figure 2.** Study workflow demonstrating the allocation of health care workers (HCWs) into 2 groups and subsequent measured outcomes. Image created with BioRender [40]. AI: artificial intelligence.




## Outcomes and Measures

Three custom instruments were administered: (1) a 10-item multiple-choice postoperative knowledge assessment quiz (score range 0 - 10); (2) a user experience survey with seven 5-point Likert scale items measuring clarity, usefulness, engagement, confidence, retention, preference, and effectiveness; and (3) an engagement metrics questionnaire capturing time spent, materials reviewed, and 3 open-ended questions. Demographics collected included age, sex, education, postoperative care experience, and AI familiarity. Complete instruments are provided in Multimedia Appendix 1. The knowledge assessment quiz was developed by the study team, informed by publicly available postoperative care resources from the American College of Surgeons [26] and Mayo Clinic patient education materials [27-36]. Content validity was established through review by 3 research fellows holding a doctor of medicine or bachelor of medicine and bachelor of surgery degree under the

guidance of a plastic surgeon. Educational materials were written at an eighth- to ninth-grade reading level, consistent with recommendations from the National Institutes of Health that patient education materials do not exceed the eighth-grade level [41].

Participants self-reported time spent reviewing materials. Following exposure to the educational content, participants completed the knowledge quiz and user experience survey.

Usability was operationalized through perceived usefulness ("The information provided was useful"), perceived clarity ("The instructions were clear and easy to understand"), and preference for clinical use ("I would prefer video-based instructions over text-based instructions in clinical practice").

Engagement time and material consumption were self-reported, consistent with standard educational research practices, to avoid technical barriers that automated tracking might introduce.

Qualitative responses were analyzed using a basic grounded theory approach, which is appropriate for exploratory pilot studies seeking to identify emergent themes without predetermined coding frameworks. Two investigators independently reviewed all open-ended responses and conducted open coding to identify recurring themes and sentiments (eg, "video was more personal," "text was hard to read," and "avatar speech seemed unnatural"). The investigators then met to discuss their independent findings, reconcile any discrepancies through consensus, and develop a final thematic framework organized into 3 categories: positive aspects, areas for improvement, and clinical practice preferences. Given the limited qualitative data (3 open-ended questions from 38 participants) and the exploratory pilot nature of the study, formal qualitative coding software, interrater reliability calculation, or thematic saturation assessment were not used. However, trustworthiness was enhanced through dual independent coding, consensus meetings, and inclusion of representative participant quotes to support identified themes. All qualitative data were collected and analyzed after completion of data collection.

### Statistical Analysis

Statistical analyses were performed using Python (version 3.12; Python Software Foundation) with the SciPy library. Independent samples 2-tailed *t* tests were used to compare mean scores between groups for all quantitative outcomes. The use of *t* tests for 5-point Likert scales is supported by evidence that parametric tests are robust to violations of normality assumptions, particularly with balanced group sizes [42]. A significance level of $P<.05$ (2-tailed) was used, with effect sizes interpreted as small (0.2), medium (0.5), or large (≥0.8). Effect sizes were calculated using Cohen *d*. Data are presented as mean (SD). Given the pilot nature of the study and multiple comparisons across secondary outcomes, results should be interpreted with appropriate caution.

As this was a pilot study designed to assess feasibility, a formal a priori power analysis was not conducted. The sample size (N=38; n=19 per group) was determined based on feasibility and resource constraints typical of pilot investigations. The observed effect sizes (Cohen $d$=0.45 for knowledge; Cohen $d$=0.81-1.38 for subjective outcomes) can inform sample size calculations for future definitive trials.

## Results

### Participant Characteristics

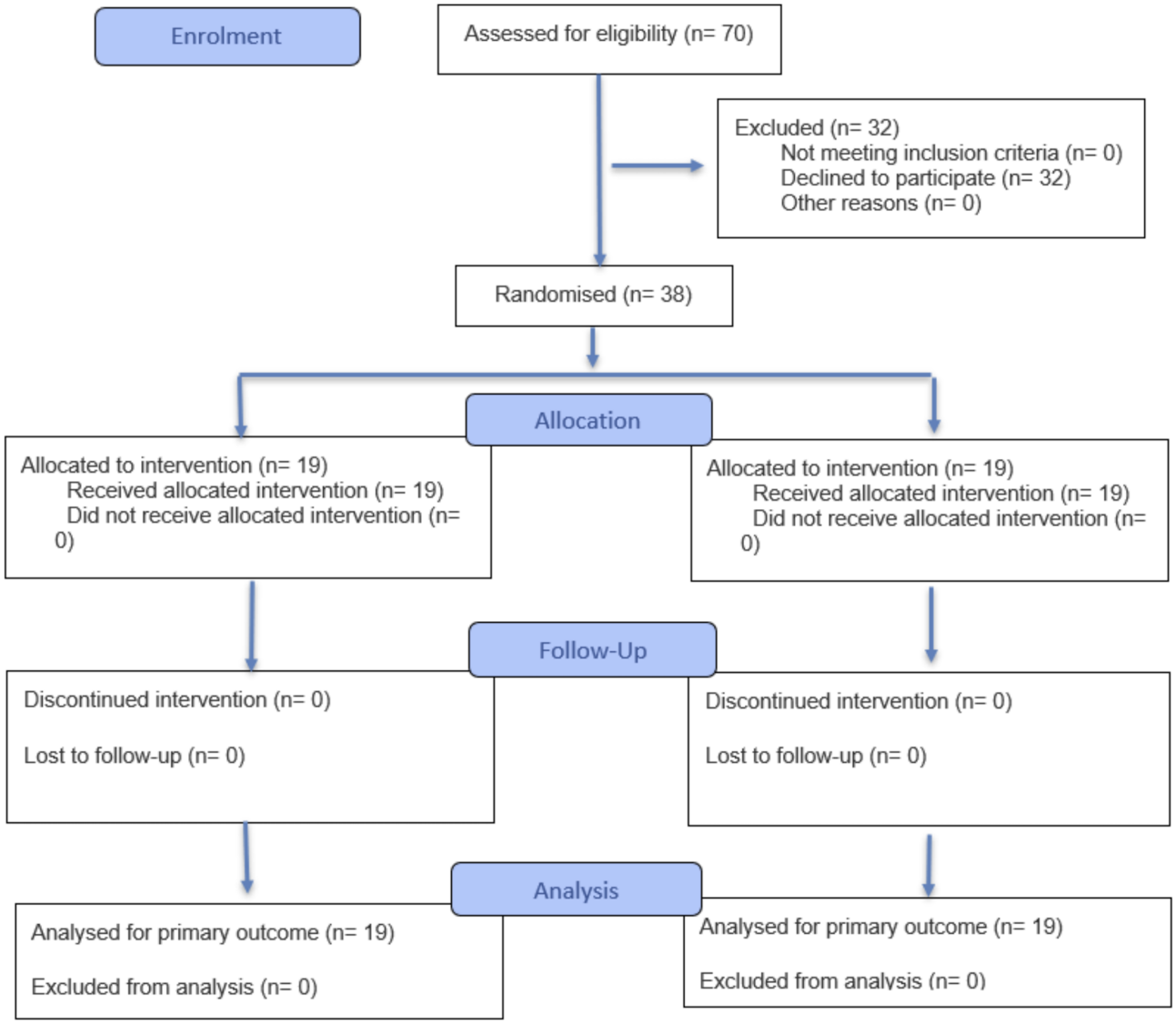
A total of 38 volunteers participated in the study, with 19 (50%) participants randomized to the text-based instruction group and 19 (50%) to the AI-generated video instruction group. Table 1 demonstrates that the text handout and AI-generated video groups exhibited similar demographic profiles. The gender distribution was nearly even across both groups, with a slight female majority in both groups. The Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth (CONSORT-eHEALTH) diagram is provided in Figure 3.

**Table .** Demographic characteristics of study participants.

| Characteristics | Text handout group (n=19) | AI[a]-generated video group (n=19) |
|---|---|---|
| Sex, n (%) | | |
| Male | 7 (37) | 8 (42) |
| Female | 12 (63) | 11 (58) |
| Age (years), mean (SD) | 32.53 (6.93) | 31.95 (7.68) |
| Educational level, n (%) | | |
| College (1-4 years) | 6 (32) | 5 (26) |
| Graduate (5-7 years) | 7 (37) | 10 (53) |
| Postgraduate (>7 years) | 6 (32) | 4 (21) |
| Previous experience with postoperative care, n (%) | | |
| Yes | 11 (58) | 11 (58) |
| No | 6 (32) | 6 (32) |
| Maybe | 2 (10) | 2 (10) |
| Familiarity with AI-generated content, n (%) | | |
| Very unfamiliar | 4 (21) | 4 (21) |
| Unfamiliar | 2 (10) | 1 (5) |
| Neutral | 4 (21) | 4 (21) |
| Familiar | 5 (26) | 4 (21) |
| Very familiar | 4 (21) | 6 (32) |

[a]AI: artificial intelligence.

**Figure 3.** Flow diagram of the progress through the phases of a randomized trial of two groups (that is, enrollment, intervention allocation, follow-up, and data analysis).




The average age of participants was comparable, with the text handout group having a mean age of 32.53 (SD 6.93) years and the AI-generated video group having a mean age of 31.95 (SD 7.68) years. Educational backgrounds were also similar, with a predominance of graduate-level education in both groups. Regarding prior experience with postoperative care, most participants reported some level of familiarity, though a notable portion, approximately one-third in each group, reported no experience. Familiarity with AI-generated content was also similarly distributed, with 29% (n=11) of participants indicating unfamiliarity, 21% (n=8) expressing neutrality, and 50% (n=19) reporting familiarity.

## Objective Knowledge Assessment

Participants in both groups completed the same 10-item knowledge quiz assessing comprehension of postoperative topics. The AI-generated video group achieved a mean score of 8.89 (SD 1.20) out of 10, while the text handout group achieved a mean score of 8.21 (SD 1.78). This difference was not statistically significant ($P$=.17; Cohen $d$=0.45). Both groups demonstrated high overall comprehension, with mean scores exceeding 80% correct.

## Engagement Levels and Time Spent

Mean completion time for the entire survey (including time spent reviewing educational materials) was approximately 25 minutes for the text handout group and 30 minutes for the AI-generated video group.

Engagement time differed significantly between the groups ($P$=.004; Cohen $d$=1.04). Participants in the AI-generated video group spent more time, on average, engaging with the instructional material than those in the text handout group. The AI-generated video group had a mean engagement time of 15.11 (SD 7.78) minutes, whereas the text handout group's mean was 8.84 (SD 4.03) minutes ($P$=.004; Cohen $d$=1.04). This large effect size (Cohen $d$>0.8) indicates a substantial practical difference in engagement time between formats. Participants in the AI-generated video group watched, on average, 7.05 (SD 2.67) videos. In contrast, participants in the text handout group read, on average, 6.68 (SD 2.92) instructions. The increased engagement time in the AI-generated video group may reflect participants pausing the videos for various reasons, including reviewing specific information, allowing personal processing time, or taking breaks.

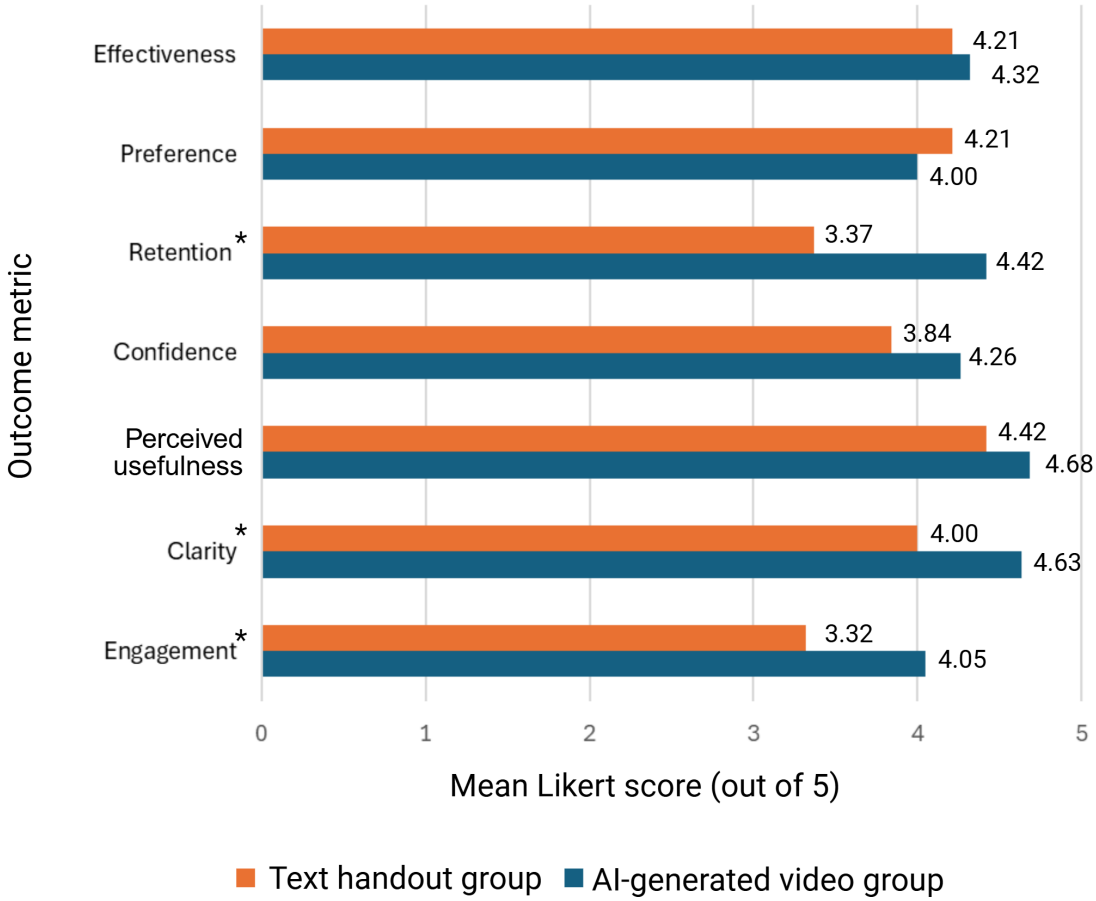
Subjective engagement ratings also differed significantly between the groups. The AI-generated video group's mean engagement rating was 4.05 (SD 0.78), while the text handout group's mean engagement rating was 3.32 (SD 1.00; *P*=.02; Cohen *d*=0.81). This large effect size suggests that participants found the video format substantially more engaging than the text format.

### Usability, Understanding, and Retention

In summary, the AI-generated video group demonstrated significantly higher engagement time, perceived clarity, subjective engagement, and information retention than the text handout group. No significant differences were observed in objective knowledge scores, perceived usefulness, confidence in understanding, preference for video-based instructions, or effectiveness in answering postoperative questions.

Participants rated the clarity of instructions significantly higher in the AI-generated video group, indicating that they found video instructions clearer than text instructions. Specifically, on a 5-point Likert scale, the AI-generated video group reported a mean clarity score of 4.63 (SD 0.50), while the text handout group reported a mean of 4.00 (SD 0.82; *P*=.007; Cohen *d*=0.93). The large effect size (Cohen *d*=0.93) demonstrates a meaningful practical difference in perceived clarity. However, participants did not perceive a significant difference in the usefulness of the information between the 2 groups. The AI-generated video group reported a mean usefulness score of 4.68 (SD 0.58), and the text handout group reported a mean of 4.42 (SD 0.61; *P*=.18; Cohen *d*=0.43). Figure 4 provides a visual comparison of all Likert scale metrics between the text and AI-generated video groups.

**Figure 4.** Comparison of text handout and artificial intelligence (AI)–generated video performance across 7 metrics, measured on 5-point Likert scales. The chart displays mean ratings for both groups. *Statistically significant differences (clarity, *P*=.007; engagement, *P*=.02; retention, *P*<.001).




Although participants in the AI-generated video group reported slightly higher confidence in understanding the instructions (mean 4.26, SD 0.65) than the text handout group (mean 3.84, SD 0.83), this difference did not reach statistical significance (*P*=.09; Cohen *d*=0.57). Participants found the video format significantly more effective for information retention. They reported a mean Likert retention score of 4.42 in the AI-generated video group, compared to 3.37 in the text handout group (*P*<.001; Cohen *d*=1.38). This very large effect size (Cohen *d*=1.38) represents one of the strongest practical differences observed in the study, indicating that participants

perceived substantially better information retention with the video format.

Both groups reported similar preferences for video-based instructions in clinical practice, with the AI-generated video group scoring a mean of 4.00 (SD 1.25) and the text handout group scoring a mean of 4.21 (SD 0.71; $P$=.53; Cohen $d$=−0.21). Finally, participants rated the effectiveness of the instructions in answering common postoperative questions similarly across both groups, with the AI-generated video group scoring a mean of 4.32 (SD 0.89) and the text handout group scoring a mean of 4.21 (SD 0.98; $P$=.73; Cohen $d$=0.12). Table 2 tabulates the engagement, usability, and retention metrics between the text handout and AI-generated video formats.

**Table .** Comparative analysis of engagement, usability, and retention outcomes between artificial intelligence (AI)–generated video and text instructional formats.

| Outcomes | AI-generated video group, mean (SD) | Text handout group, mean (SD) | *P* value |
|---|---|---|---|
| Knowledge quiz score (0-10) | 8.89 (1.20) | 8.21 (1.78) | .17 |
| Engagement time (minutes) | 15.11 (7.78) | 8.84 (4.03) | .004 |
| Subjective engagement rating (0-5) | 4.05 (0.78) | 3.32 (1.00) | .02 |
| Clarity of instructions (0-5) | 4.63 (0.50) | 4.00 (0.82) | .007 |
| Perceived usefulness of information (0-5) | 4.68 (0.58) | 4.42 (0.61) | .18 |
| Confidence in understanding (0-5) | 4.26 (0.65) | 3.84 (0.83) | .09 |
| Retention score (0-5) | 4.42 (0.84) | 3.37 (0.68) | <.001 |
| Preference in clinical practice (0-5) | 4.00 (1.25) | 4.21 (0.71) | .53 |
| Number of videos or instructions | 7.05 (2.67) | 6.68 (2.92) | N/A[a] |
| Effectiveness in answering questions (0-5) | 4.32 (0.89) | 4.21 (0.98) | .73 |

[a]N/A: not applicable. A P value was not calculated because videos and written instructions are different units of content and cannot be directly compared.

The large effect sizes observed for engagement, clarity, and retention outcomes indicate that the differences between video and text formats represent substantial practical significance beyond statistical significance alone. The moderate effect size for confidence in understanding (Cohen $d$=0.57) suggests a meaningful trend despite not reaching statistical significance ($P$=.09), likely due to our small sample size. In contrast, usefulness, clinical practice preference, and effectiveness in answering questions showed minimal practical differences between formats.

## Qualitative Feedback

Participants provided qualitative feedback on clarity, engagement, suggested improvements, clinical practice use, and format preference. Representative participant quotes illustrating these themes are provided in Table 3.

**Table .** Qualitative feedback themes by group.

| Themes | AI-generated video group | Text handout group |
|---|---|---|
| Clarity | “Crystal clear instructions” and “Clear and direct communication” | “Crisp and clear instructions” and “Clear and easy, no issues” |
| Engagement | “Engaging like a real guide” and “Lifelike, memorable instructions” | “Quick,” “Not engaging,” and “Less engaging if someone doesn’t like reading” |
| Suggested improvements | “Gestures not synced with voice,” “Unnatural facial expressions,” and “Slightly robotic voice” | “Could include diagrams or video,” “More visuals or a short summary,” and “Add an FAQ [frequently asked question]." |
| Clinical practice use | 13 out of 19 participants would use; “We are fully engaging when watching videos” | 10 out of 19 participants would use; “I’d still prefer a short video to keep people more engaged” |
| Format preference | “Engaging,” “Easy-to-follow narration,” and “Effective AI explanations” | “Straight to the point” and “Easy to skim”; many preferred video as a supplement |

## Discussion

### Principal Findings

This pilot study assessed the feasibility and user experience of AI-generated avatar videos for postoperative instruction delivery among health care workers. The AI-generated video group demonstrated significantly higher engagement, perceived clarity, and information retention than the text handout group, while objective knowledge scores did not differ significantly between groups. These findings suggest that AI-generated avatar videos are feasible to produce using commercially available platforms and may offer meaningful advantages in learner engagement, although their impact on objective knowledge acquisition requires further investigation in patient populations.

The significantly higher engagement with AI-generated videos aligns with media richness theory [42], which suggests that richer media formats are more effective in maintaining attention. The increased engagement time may reflect participants pausing or revisiting video segments to reinforce key points. The superior clarity ratings for video instructions may reflect reduced cognitive load through combined visual and auditory channels [43]. Cognitive load theory suggests that well-designed multimedia can prevent working memory overload, and the conversational tone of AI-generated avatars may create a more approachable instructional style than formal text [13,44].

Participants perceived greater retention with video instructions, consistent with dual-coding theory [45], which posits that information presented in both visual and auditory formats results in more robust encoding. A hybrid approach combining text clarity with video engagement benefits may offer the greatest educational impact [31,32]. The discrepancy between perceived and actual learning outcomes is consistent with existing research [46], suggesting that engagement benefits may not directly translate to measurable knowledge gains in populations with high baseline comprehension.

Qualitative feedback reinforced the quantitative findings, with participants noting the engaging and personal qualities of avatar-led instruction while identifying areas for technological improvement, particularly in avatar speech naturalness and gesture synchronization. These insights will guide iterative refinement of the AI-generated video format for patient-facing deployment.

Several recent studies support the growing viability of AI-generated avatar–based education in health care. Kim et al [47] found that 64% of participants preferred an AI-generated video bot over a text-based chatbot for surgical education. Coleman et al [48] demonstrated that AI-generated digital clinicians achieved significantly higher knowledge scores in a randomized controlled trial. Artsi et al [49] found that while evidence remains sparse, studies consistently report engagement benefits with AI-generated video content. Our findings align with this emerging literature, supporting the engagement advantages of AI-generated video while highlighting the need for validated knowledge assessments in future studies.

### Limitations

Several limitations of our study must be considered when interpreting the results. First, this was a pilot study with a relatively small sample size (N=38), limiting statistical power and the generalizability of our findings to broader patient populations. Participants consisted exclusively of health care worker volunteers (faculty, staff, researchers, and students), and specific clinical roles (eg, physicians, nurses, and administrative staff) were not systematically recorded. Health care workers likely possess higher baseline knowledge about postoperative care than typical surgical patients, which may explain the high quiz scores in both groups (>80%) and the lack of a significant difference between formats. This ceiling effect may have obscured the potential benefits of video instruction that could emerge in populations with lower health literacy. Future studies should include patients with diverse educational backgrounds and health literacy levels.

Second, we used a convenience sampling approach within a single institution, which may have introduced selection bias and further limited the applicability of our results to diverse clinical settings or populations. Additionally, the study evaluated only short-term comprehension and perceived retention, without assessing long-term retention, behavioral adherence, or clinical outcomes such as postoperative recovery or complication rates.

Third, many outcomes, including engagement time, retention, and clarity, were based on self-reported data, introducing the possibility of recall, reporting, or social desirability biases. Objective measures beyond immediate quiz performance were not included, limiting conclusions about true knowledge retention or clinical impact. The observed benefits in engagement and perceived effectiveness may have been influenced by the novelty effect of interacting with AI-generated avatars, which could diminish over repeated exposures or extended use. The longer engagement times observed in the AI-generated video group may have multiple explanations beyond genuine engagement, including time spent adapting to the video interface, pausing to process unfamiliar avatar presentations, or technical factors such as video buffering or lag reported by some participants. These confounding factors limit the interpretation of engagement time as a pure measure of educational engagement. The study also evaluated only a single generative AI platform (HeyGen), and results might differ substantially using alternative AI technologies or avatar presentations.

Moreover, participants reviewed educational materials outside a clinical context, potentially impacting their engagement and motivation differently compared with how real postoperative patients would experience. Furthermore, participants' technical familiarity and comfort with AI technology varied, possibly biasing results toward those more comfortable interacting with digital or AI-generated content. Finally, our study did not formally assess participants' baseline health literacy levels, limiting insights into whether the observed improvements are equally beneficial across diverse literacy groups.

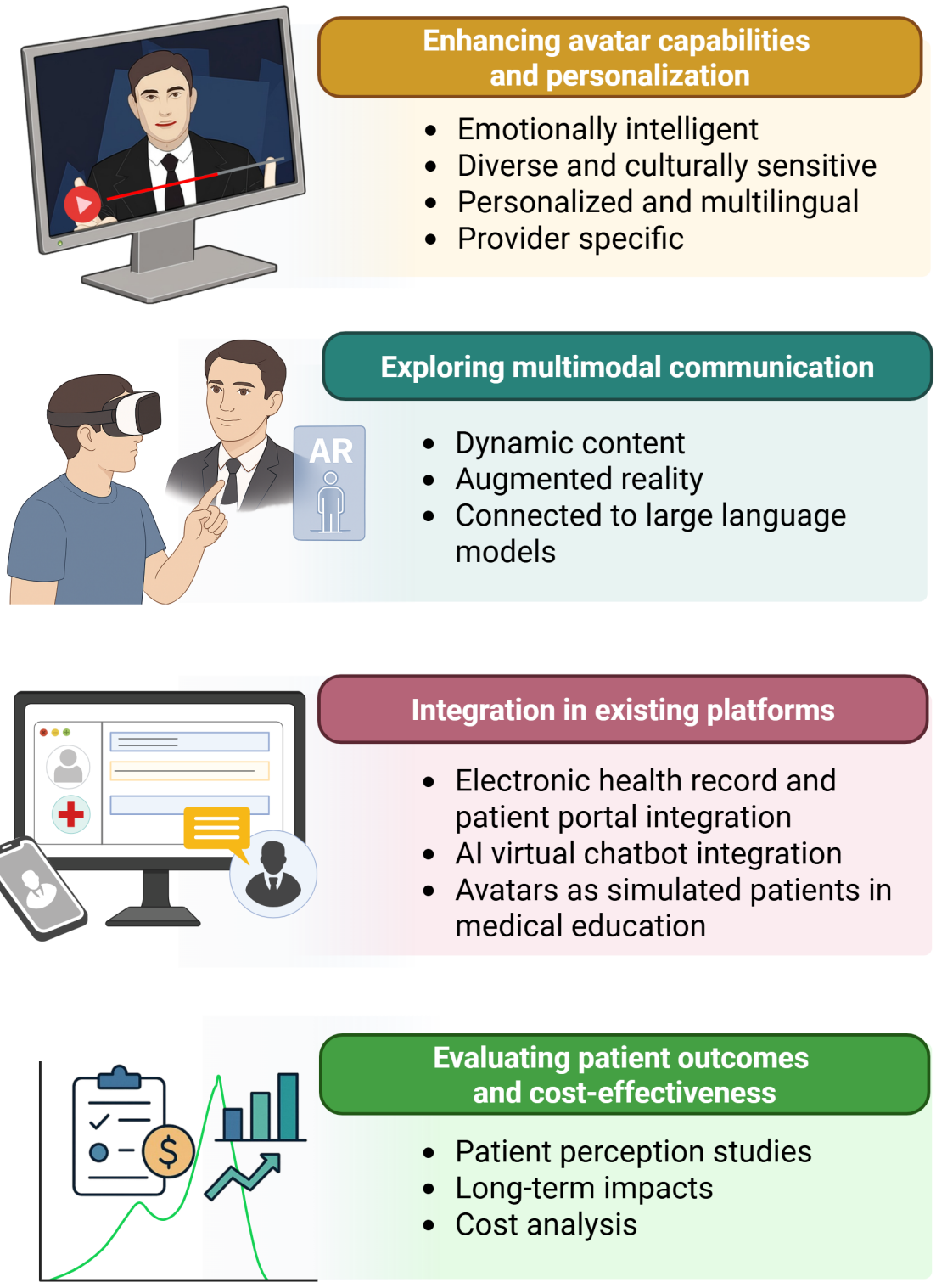
Additionally, health care workers do not approximate the general patient population: most participants had college or graduate-level education, were relatively young, and many had

prior postoperative care experience. The knowledge quiz was developed by the study team using publicly available patient education resources rather than a validated instrument, and no formal psychometric testing (eg, Cronbach α) was conducted. The multiple-choice format with easily excludable distractors may have allowed correct answers to be guessed through commonsense elimination, contributing to the ceiling effect observed in both groups (>80% correct). The reading level of the educational materials (eighth to ninth grade) exceeds the American Medical Association recommendation of sixth-grade level for general patient populations, although it is consistent with the National Institutes of Health recommendation of no higher than the eighth-grade level [41] and was appropriate for our health care worker sample.

## Future Directions

Future research should evaluate AI-generated video instructions in actual patient populations with diverse health literacy levels. Key priorities include (1) larger randomized trials in surgical patients, (2) assessment of long-term retention and clinical outcomes (eg, adherence and complications), (3) comparison across different AI platforms, and (4) cost-effectiveness analyses comparing AI-generated videos to traditional production methods. Hybrid text-video approaches and culturally tailored avatar presentations warrant investigation [24] (Figure 5 [25]).

**Figure 5.** Future directions and recommendations for the use of generative artificial intelligence (AI)–based text-to-video models in health care. Created using BioRender [50].




Future research should explore multimodal communication, including avatars with dynamic visual aids, augmented reality or virtual reality integration [51], and enhanced natural language processing capabilities through large language model integration. Longitudinal studies assessing sustained educational impact, comparative effectiveness against traditional methods, and cost-effectiveness analyses are needed. Technical priorities include electronic health record integration, 24×7 chatbot

availability, and improved accessibility features. Beyond patient education, AI-generated avatars may serve as valuable training tools for medical students in simulated clinical scenarios.

### Ethical Implications of AI

AI-generated avatars raise critical ethical concerns. Transparency is essential—patients must be informed that they are interacting with AI rather than humans to preserve trust and prevent unrealistic expectations. Accuracy requires ongoing monitoring and human oversight, as technological limitations may cause mispronunciations or misrepresentations that compromise comprehension.

Avatar-driven education should complement rather than replace human interaction, as AI cannot authentically replicate empathy and emotional responsiveness. Some users may form unintended emotional connections with AI-generated characters [52,53]; continuous availability of human support can mitigate this risk.

Digital avatars of real individuals require explicit consent regarding creation, distribution, and use. Unauthorized "deepfake" avatars pose significant risks, including misinformation, reputational damage, and erosion of trust [54,55]. Future implementations involving patient data will require robust cybersecurity and privacy protections.

Comprehensive regulatory frameworks addressing consent, ownership, transparency, privacy, accuracy standards, and accountability are essential for the responsible integration of AI-generated avatars in health care [56].

### Conclusions

This pilot study demonstrated the feasibility of rapidly producing AI-generated avatar videos for postoperative instruction using commercially available platforms. Although objective knowledge scores did not differ significantly between formats, the AI-generated video group showed significantly higher engagement, perceived clarity, and information retention. Key limitations include the small sample size, use of health care workers rather than actual patients, reliance on self-reported engagement, and use of a nonvalidated knowledge instrument with ceiling effects. Subsequent evaluation with actual surgical patients has demonstrated high trust and usability for AI physician avatars [25]. Future research should prioritize evaluation in diverse patient populations with varying health literacy levels, using validated instruments and clinical outcome measures.

## Funding

This study was supported by Mayo Clinic and the generosity of Eric Schmidt and Wendy Schmidt, Dalio Philanthropies, and Gerstner Philanthropies.

## Data Availability

The datasets generated and analyzed during this study are not publicly available due to participant privacy considerations and the terms of the institutional review board approval but are available from the corresponding author (AJF) on reasonable request.

## Authors' Contributions

Conceptualization: SAH, AJF
Data curation: SAH, AG
Formal analysis: SAH, SP
Funding acquisition: SB, CT, AJF
Investigation: SAH, SP, CAG-C, AG, BGC
Methodology: SAH, SP, CT
Project administration: JL, SB, AJF
Resources: CAG-C, JL, BGC
Software: SAH, SP
Supervision: NW, SB, CT, MAL, AJF
Validation: NW, MAL
Visualization: SAH
Writing—original draft: SAH
Writing—review and editing: SAH, SP, CAG-C, BGC, NW, SB, CT, MAL, AJF

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Study instruments: 10-item postoperative knowledge quiz, 7-item user experience survey (5-point Likert scale), and engagement metrics questionnaire administered to participants in both the text handout group and the artificial intelligence–generated video group.
[DOCX File, 39 KB - periop_v9i1e89277_app1.docx ]

Checklist 1 [DOCX File, 33 KB - periop_v9i1e89277_app2.docx ]

## References


1. Kindig DA, Panzer AM, Nielsen-Bohlman L, Institute of Medicine (US) Committee on Health Literacy. Health Literacy: A Prescription to End Confusion: National Academies Press; 2004.
2. van den Broek P, White MJ, Kendeou P, Carlson S. Reading between the lines: developmental and individual differences in cognitive processes in reading comprehension. In: Wagner RK, Schatschneider C, Plythian-Sence C, editors. Beyond Decoding: The Behavioral and Biological Foundations of Reading Comprehension: Guilford Press; 2009:107-123.
3. Engel KG, Heisler M, Smith DM, Robinson CH, Forman JH, Ubel PA. Patient comprehension of emergency department care and instructions: are patients aware of when they do not understand? Ann Emerg Med 2009 Apr;53(4):454-461. [doi: 10.1016/j.annemergmed.2008.05.016] [Medline: 18619710]
4. McCarthy DM, Waite KR, Curtis LM, Engel KG, Baker DW, Wolf MS. What did the doctor say? Health literacy and recall of medical instructions. Med Care 2012 Apr;50(4):277-282. [doi: 10.1097/MLR.0b013e318241e8e1] [Medline: 22411440]
5. Koivisto JM, Saarinen I, Kaipia A, et al. Patient education in relation to informational needs and postoperative complications in surgical patients. Int J Qual Health Care 2020 Apr 21;32(1):35-40. [doi: 10.1093/intqhc/mzz032] [Medline: 31016323]
6. Palumbo R. Examining the impacts of health literacy on healthcare costs. An evidence synthesis. Health Serv Manage Res 2017 Nov;30(4):197-212. [doi: 10.1177/0951484817733366] [Medline: 29034727]
7. Marwick C. Patients' lack of literacy may contribute to billions of dollars in higher hospital costs. JAMA 1997 Sep 24;278(12):971. [PMC: 9307333]
8. MacLeod S, Musich S, Gulyas S, et al. The impact of inadequate health literacy on patient satisfaction, healthcare utilization, and expenditures among older adults. Geriatr Nurs 2017;38(4):334-341. [doi: 10.1016/j.gerinurse.2016.12.003] [Medline: 28089217]
9. Card EB, Morales CE, Ramirez JM, et al. Impact of illustrated postoperative instructions on knowledge and retention during a cleft lip and palate surgical mission. Cleft Palate Craniofac J 2023 Aug;60(8):1010-1020. [doi: 10.1177/10556656221100052] [Medline: 35711155]
10. Fan KL, Black CK, DeFazio MV, Luvisa K, Camden R, Song DH. Bridging the knowledge gap: an examination of the ideal postoperative autologous breast reconstruction educational material with A/B testing. Plast Reconstr Surg 2020 Jan;145(1):258-266. [doi: 10.1097/PRS.0000000000006373] [Medline: 31609288]
11. Horstman MJ, Mills WL, Herman LI, et al. Patient experience with discharge instructions in postdischarge recovery: a qualitative study. BMJ Open 2017 Feb 22;7(2):e014842. [doi: 10.1136/bmjopen-2016-014842] [Medline: 28228448]
12. Krontoft AS. How do patients prefer to receive patient education material about treatment, diagnosis and procedures? —A survey study of patients preferences regarding forms of patient education materials; leaflets, podcasts, and video. Open J Nurs 2021;11(10):809-827. [doi: 10.4236/ojn.2021.1110068]
13. Katt BM, Tawfik AM, Imbergamo C, McEntee R, Beredjiklian PK, Ilyas AM. Patient comprehension of operative instructions with a paper handout versus a video: a prospective randomized control trial. J Hand Surg Am 2023 Mar;48(3):311. [doi: 10.1016/j.jhsa.2021.10.017] [Medline: 35012796]
14. O'Callaghan J. How OpenAI's text-to-video tool Sora could change science - and society. Nature 2024 Mar;627(8004):475-476. [doi: 10.1038/d41586-024-00661-0] [Medline: 38472485]
15. Li W, Yang J, Shao Q, Wu H, Li C. From text to video: what will OpenAI's Sora bring to the oncologic field? Int J Surg 2025 Jan 1;111(1):1666-1668. [doi: 10.1097/JS9.0000000000001988] [Medline: 39093893]
16. Mohamed AA, Lucke-Wold B. Text-to-video generative artificial intelligence: Sora in neurosurgery. Neurosurg Rev 2024 Jun 13;47(1):272. [doi: 10.1007/s10143-024-02514-w] [Medline: 38867134]
17. Kang Y, Wang S, Zhu L. Text-to-video models and Sora in plastic surgery: pearls, pitfalls, and prospectives. Aesthetic Plast Surg 2025 Jun;49(11):3224-3228. [doi: 10.1007/s00266-024-04583-0] [Medline: 39663218]
18. Mayer RE. The science of instruction: determining what works in multimedia learning. In: Multimedia Learning, 2nd edition: Cambridge University Press; 2009:28-56. [doi: 10.1017/CBO9780511811678.004]
19. Stull A, Fiorella L, Similuk R, Ibonie S, Mayer R. Effects of instructor presence in video lectures: rapport, attention, and learning. Presented at: Proceedings of the Annual Meeting of the Cognitive Science Society; Jul 26-29, 2023 URL: https://escholarship.org/uc/item/8033r3wc [accessed 2026-05-05]
20. Polat H. Instructors' presence in instructional videos: a systematic review. Educ Inf Technol (Dordr) 2022 Dec 23:1-33. [doi: 10.1007/s10639-022-11532-4] [Medline: 36589519]
21. Zhang L, Gallagher R, Du H, et al. Evaluate the effect of virtual nurse-guided discharge education app on disease knowledge and symptom response in patients following coronary events. Int J Med Inform 2025 Apr;196:105818. [doi: 10.1016/j.ijmedinf.2025.105818] [Medline: 39919533]
22. Wonggom P, Kourbelis C, Newman P, Du H, Clark RA. Effectiveness of avatar-based technology in patient education for improving chronic disease knowledge and self-care behavior: a systematic review. JBI Database System Rev Implement Rep 2019 Jun;17(6):1101-1129. [doi: 10.11124/JBISRIR-2017-003905] [Medline: 31021975]
23. Bedra M, Wick E, Brotman D, Finkelstein J. Avatar-based interactive ileostomy education in hospitalized patients. Stud Health Technol Inform 2013;190:83-85. [Medline: 23823383]

24. Canidate S, Hart M. The use of avatar counseling for HIV/AIDS health education: the examination of self-identity in avatar preferences. J Med Internet Res 2017 Dec 1;19(12):e365. [doi: 10.2196/jmir.6740] [Medline: 29196281]
25. Haider SA, Prabha S, Gomez-Cabello CA, et al. Artificial intelligence physician avatars for patient education: a pilot study. J Clin Med 2025 Dec 4;14(23):8595. [doi: 10.3390/jcm14238595] [Medline: 41375898]
26. Recovery and discharge. American College of Surgeons. URL: https://www.facs.org/for-patients/recovering-from-surgery [accessed 2025-11-05]
27. What to do — and avoid — before anesthesia. Mayo Clinic. 2023. URL: https://mcpress.mayoclinic.org/living-well/what-to-do-and-avoid-before-anesthesia/ [accessed 2026-03-26]
28. General anesthesia. Mayo Clinic. URL: https://www.mayoclinic.org/tests-procedures/anesthesia/about/pac-20384568 [accessed 2026-03-26]
29. Nausea and vomiting. Mayo Clinic. URL: https://www.mayoclinic.org/symptoms/nausea/basics/definition/sym-20050736 [accessed 2026-03-26]
30. Pain medicines after surgery. Mayo Clinic. URL: https://www.mayoclinic.org/tests-procedures/anesthesia/in-depth/pain-medications/art-20046452 [accessed 2026-03-26]
31. Finding the balance: opioids and pain control after surgery. Mayo Clinic. 2023. URL: https://newsnetwork.mayoclinic.org/discussion/finding-the-balance-opioids-and-pain-control-after-surgery/ [accessed 2026-03-26]
32. Cancer surgery: physically removing cancer. Mayo Clinic. 2026. URL: https://www.mayoclinic.org/diseases-conditions/cancer/in-depth/cancer-surgery/art-20044171 [accessed 2026-03-26]
33. Sparks D. Saving lives by reducing surgical site infections. Mayo Clinic. 2012. URL: https://newsnetwork.mayoclinic.org/discussion/saving-lives-by-reducing-surgical-site-infections/ [accessed 2026-03-26]
34. Cuts and scrapes: first aid. Mayo Clinic. 2024. URL: https://www.mayoclinic.org/first-aid/first-aid-cuts/basics/art-20056711 [accessed 2026-03-26]
35. Fever treatment: quick guide to treating a fever. Mayo Clinic. 2025. URL: https://www.mayoclinic.org/diseases-conditions/fever/in-depth/fever/art-20050997 [accessed 2026-03-26]
36. Bishop S. Wound location first thing to consider when deciding if medical attention is needed. Mayo Clinic. 2013. URL: https://newsnetwork.mayoclinic.org/discussion/wound-location-first-thing-to-consider-when-deciding-if-medical-attention-is-needed/ [accessed 2026-03-26]
37. Flesch R. A new readability yardstick. J Appl Psychol 1948 Jun;32(3):221-233. [doi: 10.1037/h0057532] [Medline: 18867058]
38. HeyGen. URL: https://www.heygen.com [accessed 2025-03-04]
39. BioRender licence for Figure 1. BioRender. URL: https://BioRender.com/umhj87z [accessed 2026-05-15]
40. BioRender licence for Figure 2. BioRender. URL: https://BioRender.com/iuwpq8k [accessed 2026-05-15]
41. Rooney MK, Santiago G, Perni S, et al. Readability of patient education materials from high-impact medical journals: a 20-year analysis. J Patient Exp 2021;8:2374373521998847. [doi: 10.1177/2374373521998847] [Medline: 34179407]
42. Norman G. Likert scales, levels of measurement and the “laws” of statistics. Adv Health Sci Educ Theory Pract 2010 Dec;15(5):625-632. [doi: 10.1007/s10459-010-9222-y] [Medline: 20146096]
43. Young JQ, Sewell JL. Applying cognitive load theory to medical education: construct and measurement challenges. Perspect Med Educ 2015 Jun;4(3):107-109. [doi: 10.1007/s40037-015-0193-9] [Medline: 26016430]
44. Bowers N, Eisenberg E, Montbriand J, Jaskolka J, Roche-Nagle G. Using a multimedia presentation to improve patient understanding and satisfaction with informed consent for minimally invasive vascular procedures. Surgeon 2017 Feb;15(1):7-11. [doi: 10.1016/j.surge.2015.09.001] [Medline: 26464072]
45. Kanellopoulou C, Kermanidis KL, Giannakoulopoulos A. The dual-coding and multimedia learning theories: film subtitles as a vocabulary teaching tool. Educ Sci 2019;9(3):210. [doi: 10.3390/educsci9030210]
46. Persky AM, Lee E, Schlesselman LS. Perception of learning versus performance as outcome measures of educational research. Am J Pharm Educ 2020 Jul;84(7):ajpe7782. [doi: 10.5688/ajpe7782] [Medline: 32773832]
47. Kim TS, Yu CT, Hinson C, et al. ChatGPT virtual assistant for breast reconstruction: assessing preferences for a traditional chatbot versus a human AI VideoBot. Plast Reconstr Surg Glob Open 2024 Oct;12(10):e6202. [doi: 10.1097/GOX.0000000000006202] [Medline: 39359701]
48. Coleman S, Lynch C, Worlikar H, et al. “Digital clinicians” performing obesity medication self-injection education: feasibility randomized controlled trial. JMIR Diabetes 2025 Jul 30;10:e63503. [doi: 10.2196/63503] [Medline: 40737494]
49. Artsi Y, Sorin V, Glicksberg B, et al. AI-generated videos in medical education: systematic review. BMJ Open Qual 2025 Dec 22;14(4):e003704. [doi: 10.1136/bmjoq-2025-003704] [Medline: 41429451]
50. BioRender licence for Figure 5. BioRender. URL: https://BioRender.com/yav6yeq [accessed 2026-05-15]
51. O’Connor S. Virtual reality and avatars in health care. Clin Nurs Res 2019 Jun;28(5):523-528. [doi: 10.1177/1054773819845824] [Medline: 31064283]
52. Zomorodi M, Monteleone K, Meshkinpour S, Cloutier M. MIT sociologist Sherry Turkle on the psychological impacts of bot relationships. NPR. 2024. URL: https://www.npr.org/2024/08/02/g-s1-14793/mit-sociologist-sherry-turkle-on-the-psychological-impacts-of-bot-relationships [accessed 2026-05-06]

53. Sahota N. How AI companions are redefining human relationships in the digital age. Forbes. URL: https://www.forbes.com/sites/neilsahota/2024/07/18/how-ai-companions-are-redefining-human-relationships-in-the-digital-age [accessed 2025-03-04]
54. Navarro Martínez O, Fernández-García D, Cuartero Monteagudo N, Forero-Rincón O. Possible health benefits and risks of deepfake videos: a qualitative study in nursing students. Nurs Rep 2024 Oct 3;14(4):2746-2757. [doi: 10.3390/nursrep14040203] [Medline: 39449440]
55. Westerlund M. The emergence of deepfake technology: a review. Technol Innov Manag Rev 2019;9:40-53.
56. Hendrickson L. Protecting public figures and artists' likeness in the age of AI. Identity. URL: https://www.identity.org/protecting-public-figures-and-artists-likeness-in-the-age-of-ai/ [accessed 2025-03-04]

## Abbreviations

**AI:** artificial intelligence
**CONSORT-eHEALTH:** Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth

*Edited by N Rohatgi; submitted 09.Dec.2025; peer-reviewed by D Ginting, S Mohan, X Wu; revised version received 27.Mar.2026; accepted 20.Apr.2026; published 27.May.2026.*

*Please cite as:*
*Haider SA, Prabha S, Gomez-Cabello CA, Genovese A, Collaco BG, Wood N, London J, Bagaria S, Lifson MA, Tao C, Forte AJ*
*AI-Generated Avatar Videos for Postoperative Patient Education Among Health Care Workers: Pilot Randomized Controlled Trial*
*JMIR Perioper Med 2026;9:e89277*
*URL: https://periop.jmir.org/2026/1/e89277*
*doi:10.2196/89277*

# Enhancing the User Experience of a Perioperative Digital Health Tool for Information Exchange Using a Human-Centered Design Thinking Approach: Qualitative Observational Study

Charlé Steyl[1], MBChB, DCH, DA, MMed; Carljohan Orre[2], MSc, PhD; Greg Foster[3], MSc, PhD; Hanel Duvenage[4], MSc; Michelle S Chew[5,6], MBBS, PhD; Hyla Louise Kluyts[1], MBChB, MMed, DMed

[1]Department of Anaesthesiology, School of Medicine, Sefako Makgatho Health Sciences University, Pretoria, South Africa

[2]Department of Computer Science and Media Technology, Malmö University, Nordenskiöldsgatan 1, Malmö, Sweden

[3]Department of Information Systems, Rhodes University, Makhanda, South Africa

[4]Safe Surgery South Africa, Johannesburg, South Africa

[5]Department of Anaesthesiology and Intensive Care, Linköping University, Linkoping, Sweden

[6]Department of Perioperative Medicine and Intensive Care, Karolinska University Hospital, Stockholm, Sweden

**Corresponding Author:**
Carljohan Orre, MSc, PhD
Department of Computer Science and Media Technology, Malmö University, Nordenskiöldsgatan 1, Malmö, Sweden

## Abstract

**Background:** Perioperative patient-reported outcomes (PROs) allow patients to share their experiences of surgical procedures with their health care teams using standardized measures. Despite increasing recognition of their value, PROs are not routinely used in clinical practice, partly due to limited evidence of their impact on traditional clinical outcomes and uncertainty among clinicians about their use. Digital health tools offer a promising way to integrate PROs into clinical workflows and enhance patient-clinician interaction, but their success depends on person-centered design to ensure usability and relevance. Safe Surgery South Africa, a nonprofit organization, developed the Perioperative Shared Health Record (PSHR), a secure web-based tool that enables patients to share personal health information and PROs with their anesthetist and surgeon before and after surgery. Initial implementation revealed significant user experience challenges, which contributed to poor uptake.

**Objective:** This study aimed to explore factors influencing the PSHR user experience in a low- and middle-income country (LMIC) using human-centered design principles.

**Methods:** This observational qualitative user experience study followed the 5 design thinking stages: empathize, define, ideate, prototype, and test. Semistructured interviews were conducted with postoperative patients from both the public and private health care sectors, including those with and with no prior experience using the PSHR. Thematic analysis followed the 6-phase framework described by Braun and Clarke and was structured using Karagianni's Optimized Honeycomb user experience model. A problem statement was developed, followed by ideation to explore solutions. Paper prototypes were created, refined, and tested through observation, interviews, and validated usability questionnaires.

**Results:** In the *empathize* stage, 22 interviews were conducted in the private and public health care sectors in South Africa; 7 participants had previous experience using the PSHR. In the *define* stage, participants emphasized the need for connection, feedback, information, and support through their surgical journey. Contrary to expectations, patients were not discouraged by the length of questionnaires if they perceived them as purposeful. In the *ideate* stage, the team considered user expectations and PSHR integration into care processes. In the *prototype* stage, low-fidelity mock-ups were created and refined into paper prototypes. In the *test* stage, testing with 5 participants highlighted the importance of trust, communication, and user-friendly interfaces. Feedback loops and clinician engagement were identified as key motivators for sustained use. The mean usability questionnaire scores indicated excellent usability and high levels of user satisfaction across most domains.

**Conclusions:** This study is one of the first to apply human-centered design principles to a perioperative digital health tool in an LMIC setting, addressing usability challenges and patient engagement. Key user experience factors influencing patient engagement included communication, feedback, and access to information throughout the surgical journey. Digital health tools such as the PSHR can strengthen communication and support person-centered perioperative care by integrating PROs into clinical workflows and care processes.

***(JMIR Perioper Med 2026;9:e79349)*** doi:10.2196/79349

**KEYWORDS**

patient-reported outcome measures; person-centered care; patient participation; digital health; perioperative care; user-centered design; user experience research; human-centered design; universal design; design thinking

## Introduction

Patients presenting for surgical procedures often feel vulnerable and may become overwhelmed by information that lies outside their usual frame of reference. Many also experience significant physical and emotional symptoms before and after their operation [1,2]. Time constraints and brief interactions during busy ward rounds can limit opportunities for patients to voice concerns or seek clarification [1,3]. In this context, there is a risk that patients feel depersonalized: reduced to passive participants in a system rather than active participants in their own treatment [1]. Furthermore, perioperative clinicians such as surgeons and anesthesiologists often prioritize traditional problem-focused postoperative outcomes such as morbidity and mortality rates, which do not necessarily reflect the outcomes that matter most to individual patients [4-6].

Person-centered care addresses these challenges by recognizing the individual behind the patient: a human being with values, emotions, and goals, and by fostering a partnership that supports patient autonomy and active participation in care decisions [7,8]. Evidence suggests that patients who are empowered and engaged in their own health care may have better outcomes [3,9-11]. Perioperative patient-reported outcomes (PROs) provide one means of achieving this by allowing patients to communicate their experiences of surgical procedures using standardized questionnaires, known as patient-reported outcome and experience measures (PROMs and PREMs) [2,4,12-14]. Various PROs have been defined in the perioperative sphere, including patient satisfaction, quality of recovery (a short-term outcome after surgery), and quality of life (a longer-term outcome after surgery) [14]. Perioperative PROMs and PREMs give individual patients a means to communicate how they are recovering after surgery in such a way that it can be compared between patient groups and procedure types [5,14]. The data can be used to track quality of care over time [6].

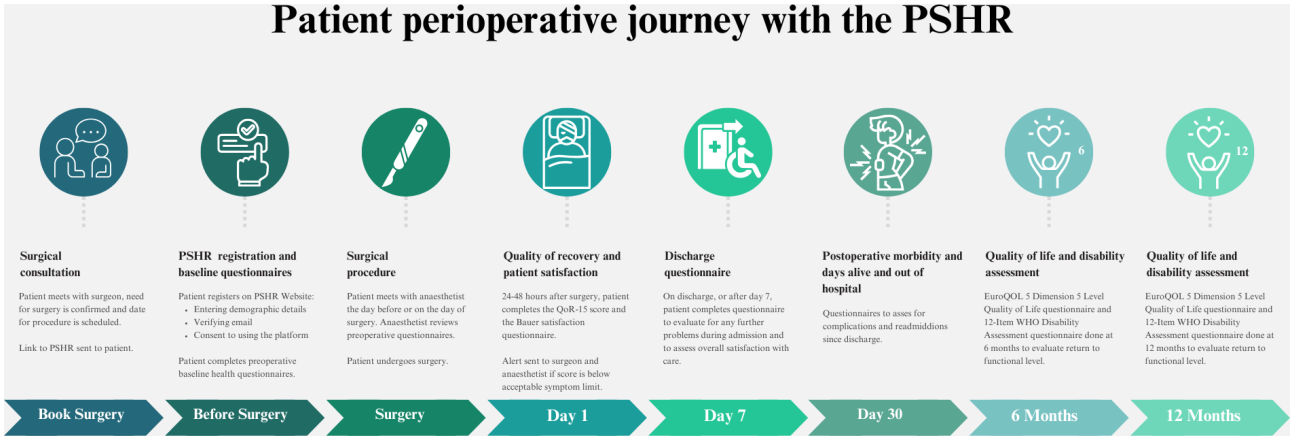
Consensus guidelines recommend the use of PROMs and PREMs in clinical research, and their implementation in perioperative care is increasingly studied [2,6,15-18]. In daily practice, however, the routine use of PROs is hampered by their time-intensive nature, limited evidence linking them to traditional outcomes such as complications or mortality, and uncertainty among clinicians and patients regarding their value [2,6,12,16,17]. These challenges may reflect insufficient person-centeredness in the application of PROs and a lack of responsiveness of health care teams to the information provided by patients [2,19].

Digital health tools offer a promising way to integrate PROs into clinical workflows and to enhance communication between patients and clinicians [17,18,20]. Achieving this, however, requires a human-centered design approach to ensure that digital tools are responsive to diverse user needs and care contexts [21]. Human-centered design forms part of the broader design thinking framework: an empathetic, iterative process that involves end users throughout development to create tools that are understandable, useful, and enjoyable to use [22-25]. Applying these principles through user experience research, which uses interviews, surveys, and usability testing to explore how people interact with digital systems, helps developers create tools that are more intuitive, engaging, and relevant to real-world care [21,26]. Learning from established digital health platforms and implementation of electronic medical record systems can help create digital health tools that support care across the patient journey [27-31].

Perioperative digital health tools have shown promise in high-income countries (HICs) [17,18], but their implementation and use remain underexplored in low- and middle-income countries (LMICs). In South Africa, barriers to large-scale adoption of digital health solutions include limited digital literacy and unequal access to technology and internet across social, economic, and geographic groups [32,33]. These disparities reflect broader inequalities within the health system, where a tax-funded public sector provides care for most of the population but is underresourced, while a private sector funded through medical schemes and out-of-pocket payments serves a minority yet absorbs a large share of resources [34-36]. The public sector continues to rely largely on paper-based documentation, with uneven implementation of systems to capture routine health information and limited electronic record keeping compared with the private sector [37-41]. The private sector is data-rich, with more electronic data systems, but its datasets are typically siloed and not routinely accessible to public governance systems, clinicians, or patients [39,42]. Neither sector currently supports routine or large-scale capture of PROs, limiting opportunities to measure and improve perioperative care from the patient perspective. Addressing these challenges requires context-specific digital solutions that can be designed to strengthen perioperative care in South Africa.

In response to these challenges, Safe Surgery South Africa [43], a research-driven nonprofit organization, developed the Perioperative Shared Health Record (PSHR) [44], a web-based digital tool enabling patients to share baseline preoperative data and postoperative PROs with their surgeon and their anesthetist for up to a year after surgery. Preoperative data can be used in risk stratification and shared decision-making, whereas postoperative PRO data can improve patient care. Data are stored on a secure server but are accessible to both patient and clinician. The system was designed to function across the public and private health care sectors to promote broader accessibility. Figure 1 describes the use of the PSHR in capturing the perioperative journey of a surgical patient.

**Figure 1.** Patient journey when using the PSHR. EuroQOL: European Quality of Life questionnaire; PSHR: Perioperative Shared Health Record; QoR-15: 15-Item Quality-of-Recovery questionnaire; WHO: World Health Organization.




Initial use of the PSHR in the private health care sector, during the South African Collaborative Surgical Outcomes Study (SACSOS; ClinicalTrials.gov NCT05052021), identified numerous user experience challenges that led to low patient and clinician engagement, which reduced the effectiveness of the PSHR. The registration process was cumbersome, requiring active support for patient users to complete it. In addition, some questionnaires were perceived as lengthy and burdensome, potentially discouraging patients from completing subsequent assessments. As many of the questionnaires are standardized tools designed for specific purposes, content modifications were not always feasible.

The aim of this observational study was to determine the factors that influence the patient user experience of the PSHR as a tool to support perioperative care. The primary objective was to evaluate the user experience of patients in South Africa who had previously used the PSHR during SACSOS. Further objectives were to explore the user needs of patients who have no prior experience with the PSHR and to gain deeper insights into the future design requirements of the PSHR.

To achieve this aim, this study used a human-centered design thinking approach, which required a multidisciplinary team that could combine clinical insight, technical expertise, and practical experience. In this study, anesthesiologists contributed their understanding of perioperative workflows and patient-clinician communication. The information systems, computer science, and media technology researchers applied user experience and design thinking principles to translate patient needs into feasible design solutions. One of the anesthesiologists (CS), with research expertise in PROs, and one of the information systems researchers (CJO), with personal experience as both a patient and a hospital representative, brought perspectives that ensured that the patient remained central throughout the course of the project. The researchers brought together expertise from South Africa and Sweden, combining experience with emerging and advanced digital health systems and perspectives from LMIC and HIC settings. These efforts aim to improve the PSHR's usability and provide insights for more person-centered digital health design.

## Methods

### Ethical Considerations

This observational qualitative user experience study was approved by the Sefako Makgatho Health Sciences University Research Ethics Committee on November 16, 2023 (SMUREC/M/513/2023:IR), and registered on the National Health Research Database on January 28, 2025 (NHRD: GP_202501_070). Written consent was obtained from all participants and patient privacy and confidentiality was respected by deidentifying patient data and removing any identifiable features from images used in publication. Participation was voluntary with no compensation paid to participants.

### Setting

The study took place in South Africa, with postoperative patients and carers recruited via purposeful sampling in both the private (insurance-funded) and public (tax-funded) health care sectors. These 2 sectors are vastly different in South Africa, with a different patient demographic and a significant difference in availability of resources.

### Research Team

The research team consisted of 4 female researchers and 2 male researchers. Three of the female researchers (CS, HK, and MC) are practicing anesthesiologists. HD has a research and development background. HK is the founder of Safe Surgery South Africa and focuses on data-driven solutions to improve perioperative risk stratification and surgical outcomes. CS has a clinical and research interest in PROs and previously recruited patients for the PSHR as part of the SACSOS study, maintaining professional relationships with these participants. One of the male researchers, CJO, comes from a digital health and media background and is a patient representative on a hospital management board after surviving cancer. CJO is also involved in an online cancer rehabilitation program spearheading the use of PROMs and PREMs to improve care processes. The other male researcher, GF, comes from a user experience and design science research background.

## Methodology

This study was informed by human-centered design principles. The design thinking process was used to structure the study around 5 phases: empathize, define, ideate, prototype, and test [23,25]. In keeping with a person-centered approach, participants informed the project from the outset by sharing their perioperative experiences and needs. These insights guided ideation and design decisions by the investigators, and participants were reengaged during prototype testing to evaluate and improve solutions based on earlier input.

### *Phase 1: Empathize*

The first phase of the project focused on empathizing with PSHR users by interviewing 2 distinct groups. The first group, recruited from the private health care sector, had prior experience using the PSHR before and after surgery through SACSOS (group 1: PSHR experience). The second group included individuals from both the public and private sectors who had undergone surgery but who had no prior exposure to the PSHR (group 2: no PSHR experience). Participants were invited to take part via email, telephone, or by in-person invitation.

For ease of reference when presenting participant quotations, each participant is assigned a letter prefix. “P” denotes participants in group 1, who had used the PSHR before, “U” denotes participants in group 2 from the public sector, and “I” denotes participants in group 2 from the private sector.

Semistructured interviews were conducted between November 2023 and January 2024. Through storytelling, empathy maps and patient journey maps were created. Demographic data were recorded in REDCap (Research Electronic Data Capture) [45] and exported to MS Excel (version 2411; Microsoft Corp). The interviews began by exploring all patients’ perioperative experiences. Group 1 participants were then asked about their experiences using the PSHR, while group 2 participants received a brief demonstration of the PSHR before discussing their expectations of a digital tool for perioperative information exchange. The interview guide is included in Multimedia Appendix 1.

Interviews were recorded and transcribed using transcription software (Transcribe—Speech to Text, version 4.20.5; DENIVIP Group LLC) on an iPad dedicated to the project. Transcriptions and audiovisual files were stored in a password-protected online folder. Transcriptions were checked for accuracy by CS and HD and reviewed by all investigators before data analysis. Transcriptions of the initial interviews were systematically coded and analyzed thematically using Nvivo (release 14.23.4(49); Lumiere) by CS and HD. The thematic analysis followed the 6-phase framework described by Braun and Clarke [46]. To explore the main themes related to participants’ experiences and expectations of the PSHR, responses were systematically coded and categorized using Karagianni’s Optimized Honeycomb model [47]. This model, commonly used in user experience research, structures the analysis of how users interact with a product by breaking down their experiences into 3 primary dimensions: Use, Feel, and Think [47-51]. By applying this framework, we identified patterns in the data and gained deeper insight into the factors influencing user experience of the PSHR.

### *Phase 2: Define*

The information obtained during phase 1 was used to create a problem statement and summary of findings.

### *Phase 3: Ideate*

Insights from the initial interviews and the defining phase informed an ideation phase, during which various solutions were brainstormed by CS, HD, HK, CJO, and GF to enhance future implementation of the PSHR and also taking into account the interoperability with electronic health records.

### *Phase 4: Prototype*

Paper prototypes for the PSHR were created in Balsamiq Wireframes for Desktop (version 4.8.1; Balsamiq Studios LLC). Paper prototypes were refined based on research team group discussions and during user testing.

### *Phase 5: Test*

User testing with the paper prototypes took place in December 2024 with patients and carers who were recruited via email, telephonically, and in person, with the aim to recruit both patients and carers who had used the PSHR before (“expert users”) to determine whether insights learned from them during initial interviews had improved their user experience. and patients and carers who had no prior experience of the PSHR (“novice users”), to determine their first time user experience with the system.

As all the expert users would be from the private health care sector, novice users were recruited from the public health care sector. To recruit the expert users, attempts were made to contact all 7 participants from group 1; 3 could not be reached at all, 1 initially agreed but later withdrew, and 3 consented and participated. For the novice users, we intentionally sought individuals with no prior exposure to the platform, including through earlier interview phases, to ensure unbiased, first-time user perspectives. This necessitated recruitment of new participants. Eligibility for user testing included being conversant in English, having basic familiarity with mobile phone and computer use and with the use of the internet. Testing was undertaken by 4 investigators (CJO, CS, GF, and HD), 1 acting as the “computer” to change paper “screens” based on user actions, 1 facilitating the scenario, and 2 observing the interaction; sessions were also audio-recorded for later analysis. For ease of reference, “T” denotes participant responses in the user testing phase.

Participants were asked to complete four tasks during the prototype testing: (1) registering and consenting, (2) completing preoperative baseline questionnaires, (3) finding additional information on the PSHR, and (4) completing postoperative quality of recovery and patient satisfaction questionnaires (Multimedia Appendix 2).

The System Usability Scale (SUS) was used to assess usability after the user prototype testing, as this is a well-established tool that has been found to have good reliability to evaluate the usability of digital systems. The SUS is a 10-statement scale

for usability of electronic health applications with good reliability (Cronbach α=0.911) and good face validity [52]. The SUS score ranges from 0 to 100 where higher scores indicate greater usability [53]. A mean SUS score of 68 (SD 12.5) represents the average benchmark for digital health apps [54].

User experience and usability aspects were assessed after prototype testing with the User Experience Questionnaire (UEQ), a well-established tool that evaluates 6 aspects including attractiveness, effectiveness, perspicuity, dependability, stimulation, and novelty with 26 pairs of terms that are scored from 1 to 7 [55]. The questionnaire has good construct validity and good reliability (Cronbach α for 5 of the 6 aspects is above 0.7) [55]. Scoring is done with a downloadable tool, with values ranging between −3 (horribly bad) and +3 (extremely good) [55]. Scores should be evaluated against current benchmarks, freely available for download [56,57].

Prototype user testing was analyzed by 4 investigators (CJO, CS, GF, and HD) who took part in the process using interviews and observation. Thematic analysis of the user tests was done based on research team discussions following the user tests. The SUS and the UEQ were scored in MS Excel according to the guidelines in their reference papers [53-55,57].

## Results

### Phase 1: Empathize

A total of 22 initial semistructured interviews were conducted as part of empathizing with users. Participant demographics are summarized in (Table 1). All the participants had access to a mobile phone; all but 2 participants in group 2 in the public sector had access to the internet on their mobile phone. Most participants (16/22, 73%) usually used their mobile phones for accessing the internet, whereas 4 out of 22 (18%) participants preferred to use a computer for internet access, and 2 out of 22 (9%) participants did not use the internet at all.

**Table .** Demographic data for semistructured interviews conducted for phase 1: empathize.

| ID[a] | Age (years) | Sex | Language | Race | Prior PSHR[b] experience | Education | Surgery |
|---|---|---|---|---|---|---|---|
| P1 | 33 | Female | English | Black | Yes | After school qualification | Major abdominal |
| P2 | 42 | Female | English | White | Yes | After school qualification | Major abdominal |
| P3 | 62 | Female | Afrikaans | White | Yes | After school qualification | Major abdominal |
| P4 | 22 | Female | English | Black | Yes | After school qualification | Major abdominal |
| P5 | 80 | Male | Afrikaans | White | Yes | After school qualification | Major abdominal |
| P6 | 72 | Male | Afrikaans | White | Yes | After school qualification | Major abdominal |
| P7 | 46 | Male | Afrikaans | White | Yes | After school qualification | N/A[c]—assisted family member |
| U1 | 34 | Female | Setswana | Black | No | Secondary school not completed | Major abdominal |
| U2 | 22 | Male | Northern Sotho | Black | No | Secondary school not completed | Major abdominal |
| U3 | 43 | Female | Tsonga | Black | No | Secondary school completed | Vascular surgery |
| U4 | 65 | Male | Setswana | Black | No | Secondary school not completed | Vascular surgery |
| U5 | 48 | Female | Afrikaans | White | No | Secondary school not completed | Bariatric surgery |
| U6 | 29 | Female | Northern Sotho | Black | No | After school qualification | Bariatric surgery |
| U7 | 42 | Female | Afrikaans | Colored[d] | No | After school qualification | Bariatric surgery |
| U8 | 39 | Female | English | Black | No | After school qualification | Bariatric surgery |
| U9 | 42 | Male | French | Black | No | Secondary school completed | Vascular surgery |
| U10 | 45 | Female | Setswana | Black | No | After school qualification | Vascular surgery |
| I1 | 64 | Female | Afrikaans | White | No | After school qualification | Orthopedic |
| I2 | 41 | Female | Afrikaans | White | No | After school qualification | Bariatric surgery |
| I3 | 54 | Female | Afrikaans | White | No | After school qualification | Breast surgery |
| I4 | 76 | Male | Afrikaans | White | No | After school qualification | Orthopedic |
| I5 | 42 | Male | Setswana | Black | No | After school qualification | Orthopedic |

[a]User ID explanation: “P” denotes participants who had used the PSHR before, “U“ denotes participants from the public sector with no prior experience

of the PSHR, and "I" denotes participants from the private sector with no prior experience of the PSHR.

[b]PSHR: Perioperative Shared Health Record.

[c]N/A: not applicable.

[d]In South Africa, the term "Colored" refers to a distinct cultural and ethnic group with mixed ancestry, recognized as a separate demographic category.

Thematic analysis of the interviews identified 3 main themes: Patient Journey (both groups), PSHR Experience (group 1), and PSHR Expectations (group 2), each with subthemes related to patient engagement and user experience. Detailed findings are provided in Multimedia Appendix 3.

### *Patient Journey*

In understanding the patient perioperative journey (main theme), the following subthemes were identified in both user groups: information-seeking behaviors, emotional response, postoperative difficulties, interaction with health care providers, and advice to other patients. Some participants actively sought more information, either by doing online searches or talking to family members or patients who have been through a similar situation.

> *I was checking [online] how long it's going to be the operation. Okay. Yeah. And how was going to be the pain? How I was cut, a lot of it.* [U1]
>
> *I go and check like, like the food I have to eat. And then the thing I didn't Google about it is the pills the most. But you know that if you want to go look, you can go, you can go find.* [U3]
>
> *So I had my sister-in-law, who is a general practitioner, check for the results and then she was the one [that told me].* [P2]
>
> *And I was also following [on social media], uh, people that will talk about their experience, you know.* [P1]
>
> *I must say the information from other patients helped a lot, knowing what someone else went through, their experiences, how they felt, what the cost implications were, how they paid it, all of that helped a lot.* [I2]

In both the private and public sectors, there were participants who indicated that they avoided looking for any additional information:

> *I don't really want to Google stuff because you always, there's always stuff. Too much information.* [P3]
>
> *I think that would've scared me off a little bit more if I knew truly what was to come.* [P4]
>
> *...because you know when you Google things you don't always get the right information. And it can be very scary.* [I3]
>
> *So, I give up to an extent that I did not even want to stress myself about the Google information. Because others there are just making some speculations.* [U10]

All the participants experienced some form of emotional turmoil in the time after their diagnosis and before they had surgery, with some describing being in denial, feeling helpless, and isolated:

> *I started like shaking and getting worried. Yes. Like now it's getting worse. And like I took it easy, like okay, fine. I went back to work instead of going to the doctor.* [I5]
>
> *But now it all became too much. It just felt like you take one step forward and like five steps back… It felt like I was in constant pain and I also felt very helpless.* [P1]
>
> *It was very... because nobody can come in with you and then you're there alone and then they don't communicate well, doctors all the time, some of them.* [P2]

One participant said an uplifting conversation with her surgeon gave her hope before her surgery and this helped her carry on with her treatment:

> *That answer, that one sentence, and with such conviction, uh, brought back my, um, my hope.* [P3]

Participants in both groups described some physical difficulties in the postoperative period:

> *I think the first two weeks were the hardest really. And the vomiting was much worse at home. Yeah. Uh, the pain also from eating was really terrible.* [P4]
>
> *But that was also the worst thing that I had the operation, because it was very painful! I didn't expect it would be so painful!* [U7]

Two participants commented that being informed and being able to contact their surgeon made the perioperative journey easier.

> *So being informed. Yeah. Makes you feel more reassured.* [P3]
>
> *The interesting thing about this surgeon's practice, that I have not come across before, is that he gives you a 24hr whatsapp number that you can use any time of the day if you have problems or questions. There is always someone that responds—that is not something that everyone would do.* [I4]

Participants in both the public and private sectors described their interaction with their health care providers in positive terms, and they valued in-person communication:

> *It made such a difference that [the anaesthetist] were there and [she] could explain to us what was going to happen, it made us feel a lot more secure and calm.* [P6]
>
> *...what I felt was more these people are taking care of me. I was positive. I could see these other people (points to other patients in ward), they are getting more healthy.* [U2]

Participants offered advice to others that reflected both practical and emotional preparation for surgery. Some emphasized the need to prepare physically by doing breathing exercises and

maintaining mobility, while others recognized the emotional impact of surgery on the patient and their families.

> *It's the emotional side of things that takes quite a toll. But not only on [the patient], but on [the family] too. So I think the emotional strain on both was tough.* [P7]

Several advised future patients to listen to their doctors, follow instructions carefully, and trust the care team. Others highlighted the importance of patience and realistic expectations, especially regarding the time needed for healing:

> *I would tell them that it is very important to listen to what the doctor tells you. To stick to the rules...And I would tell them that they shouldn't be scared to go through with it.* [I1]

> *I would explain my journey the way it is, then they can come here and get that help because it's a better help than any other.* [U3]

> *As a patient, I will say first thing first you need to be patient. You need like...healing is a mercy. It won't just happen overnight.* [I5]

### *PSHR Experience*

Group 1 included 6 patients and 1 family member who had used the PSHR. Their user experience, analyzed according to the Optimized Honeycomb model, is summarized in (Table 2), with supporting quotations in Multimedia Appendix 3.

**Table .** Codes related to experience or expectations of the Perioperative Shared Health Record[a].

| | Subthemes[b,c] | | | | | | |
|---|---|---|---|---|---|---|---|
| | Use | | | Feel | | Think | |
| | Findable | Accessible | Usable | Desirable | Credible[d] | Useful | Valuable |
| PSHR[e] experience: Group 1: PSHR exposed; private sector; 7 interviews conducted with 4 women and 3 men (2 Black and 5 White, aged 18 - 80 years) | • WhatsApp link preferred (9)<br>• Email not used frequently (5)<br>• Desktop used initially (1) | • Mobile phone access preferred (6)<br>• Device limitation (2)<br>• Postoperation no access to glasses (1)<br>• Not comfortable with technology (1) | • Easy to use (4)<br>• Technical problems/bugs (4)<br>• Login process difficult (2)<br>• Importance of feedback (2)<br>• Font size too small (1)<br>• Medical language tricky (1)<br>• Loss of interest over time (1) | • Can complete at home (2) | • No codes | • Personal connection to doctor (10)<br>• Means to feedback from doctor (8)<br>• Patients' ability to express their needs (5)<br>• Benchmarking (4)<br>• Ability to give family access to platform (1) | • Improved care (5)<br>• Patient involvement (1)<br>• Altruism (1) |
| PSHR expectations: group 2: PSHR unexposed; public sector; 10 Interviews conducted with 7 women and 3 men (7 Black, 2 White, 1 Colored[f], aged 22 - 65 years) | • No codes | • WhatsApp link preferred (12)<br>• Device limitation (4)<br>• Mobile phone access preferred (3)<br>• Email not used frequently (3)<br>• No access to phone postoperatively (2) | • No codes | • No codes | • No codes | • Communication channel (9)<br>• Means to feedback (5)<br>• Benchmarking (4)<br>• Efficiency (3) | • Improved care (2)<br>• Personal connection (2) |
| PSHR expectations: group 2: PSHR unexposed; private sector; 5 interviews conducted with 3 women and 2 men (1 Black and 4 White, aged 41 - 76 years) | • No codes | • WhatsApp or SMS (2)<br>• Email not easily accessible (2)<br>• Email link useful (1)<br>• No access to phone postoperatively (1) | • No codes | • No codes | • Negative feedback may impact care (1) | • Efficiency (3)<br>• Source of information (3)<br>• Curated list of information (1)<br>• Information about doctor (1)<br>• Communication channel (1)<br>• Link to other patients/support groups (1) | • Altruism (6) |

[a]Main theme: PSHR Experience (group 1) and PSHR Expectations (group 2).

[b]Codes from the text were sorted according to the 7 aspects of user experience from the Honeycomb Model (Findable, Accessible, Usable, Desirable,

Credible, Useful, and Valuable) [47]. Code names are included under each aspect, with the code count in parentheses.

[c]Subthemes: Use, Feel, and Think according to Karagiani's Optimized Honeycomb Model [47].

[d]Note that the aspect Credible falls under both Feel and Think.

[e]PSHR: Perioperative Shared Health Record.

[f]In South Africa, the term "Colored" refers to a distinct cultural and ethnic group with mixed ancestry, recognized as a separate demographic category.

Most participants preferred to follow WhatsApp links to find their way to the PSHR on their smartphones:

> *I do prefer that it was easy to use on my phone. So if it can be improved, it must still just be improved. Mainly for, for like a smartphone.* [P1]
>
> *Yes, because on personal email you're not visiting that often, so it gets lost with all the other stuff. So, I believe your preferred communication is WhatsApp.* [P7,]

One participant commented on the potential difficulty of using a smartphone on the first day postoperatively:

> *It was a bit difficult. Different. If you wear glasses and you don't have glasses on, and you're on morphine. But it wasn't, it wasn't impossible.* [P2]

One of the older participants indicated that they were not comfortable with technology, which is a potential barrier to using a digital tool such as the PSHR:

> *No, I'm not so comfortable with my phone. The internet on there is not something that I usually use.* [P5]

Contrary to expectations, the length of the questionnaires was not perceived in a negative light:

> *It did just go on and on and on. Um, I, I think in my head it was just all part of, just part of the process, preparation and the process you had to do, you know, and making sure that everything is fine.* [P1]
>
> *It was easy to answer. It doesn't take too long.* [P4.]

However, there was some concern about medical jargon and font size:

> *There's some, um, uh, of the wordings and stuff that I really didn't understand.* [P3]
>
> *I think the only complaint if I need to complain about improvements will be the size of the font perhaps.* [P7]

Participants valued the PSHR for enhancing their engagement and improving the quality of care they received:

> *...[the doctor] was able to quickly know and come back and improve my care, you know?* [P1]
>
> *You feel that you were more involved in the planning of your care.* [P2]

Feedback from the surgeon or anesthetist emerged as an important motivator for continued use:

> *[The anaesthetist] had read what was going on. She came and she asked what was going on, and I explained that and she worked around it and talked to the staff.* [P1]
>
> *If I didn't get feedback, I wouldn't have filled in anymore. I would've done the first one and left it at that.* [P2]

Participants appreciated being able to reflect on their recovery:

> *All of this is quite relevant because it lets you think about your own wellbeing and progress.* [P6]

Interestingly, some participants were also motivated by altruism, expressing a desire to help others:

> *...if my information can help somebody else get through a very difficult situation...then I feel it's worth it.* [P3]

Suggestions for improving the PSHR centered around information sharing and being able to contact patients who had been through a similar procedure:

> *I suppose especially for, for large operations, it might help people to know who the anaesthetist is and have like a name and a, a maybe a photograph of your doctor on the system...And maybe info about postoperative care. Because I mean all these ops have different things and I didn't know I was gonna go to need dietary requirements after the first liver operation...So having that as a portal to kind of find information may be useful.* [P2]
>
> *I think having someone else who knows, you know, what you've been through would be nice. Yeah. They can give you kind of, like a perspective on what to expect.* [P4]
>
> *I would prefer to see a video, just a more informal video and then follow up with a verbal conversation just before the operation.* [P7]

### *PSHR Expectations*

Following a brief demonstration of the PSHR, 10 public and 5 private sector postoperative patients with no prior experience of the PSHR (group 2) were interviewed about their expectations of a digital information-sharing tool. Their expectations are summarized in the second part of Table 2, with illustrative quotations provided in Multimedia Appendix 3.

Most participants indicated a preference for accessing the PSHR via a WhatsApp link on their smartphones:

> *I think overall on one's phone is just better, it is more accessible.* [I2]
>
> *We have emails, but we don't use it so much.* [U2]
>
> *Anything that is easy for you is easy for me. But really Whatsapp is easiest.* [U10]

Participants indicated that they would value features such as curated information, feedback from their surgeon or anesthetist, and the ability to track their recovery progress.

> *I don't want to get the information by doing a google search. I want information that comes from the doctor themselves, so that I know it is correct.* [I1]
>
> *You know, if I think back to my work again, clients want to be heard… and now in my setting I am not upset about anything, but it may still be nice to be acknowledged, if I fill something in, it would be nice to get a message or a call to confirm that my responses were seen.* [I3]
>
> *But when you check, keep on checking on your patient, it's good because if I feel something on me, I have to let you know. Then you'll ask me maybe then to come back at hospital. Then you can check that and sort it out.* [U3]
>
> *I mean, if they know what your baseline is, what my baseline is, how my life, my, my health is, you know, then they'll know how to proceed. With any procedure for that matter.* [U5]

Participants also indicated that they would be motivated to use a tool such as the PSHR by knowing that they would contribute data that could help others:

> *I would actually do it more for the greater good to contribute to ongoing medical knowledge and learning.* [I1]
>
> *I think I would still contribute my data if I knew it went for a good cause and if my doctor asked for it.* [I2]
>
> *If it'll help someone with the same problem that I have, it's important to share it.* [I5]

Potential barriers to using the PSHR are high costs of data and low digital confidence:

> *When I'm at home, I don't see that airtime. Because it's a cost of money. It's very expensive.* [U3]
>
> *...all the fancy phones, the internet, all that stuff, that's not for me.* [U4]

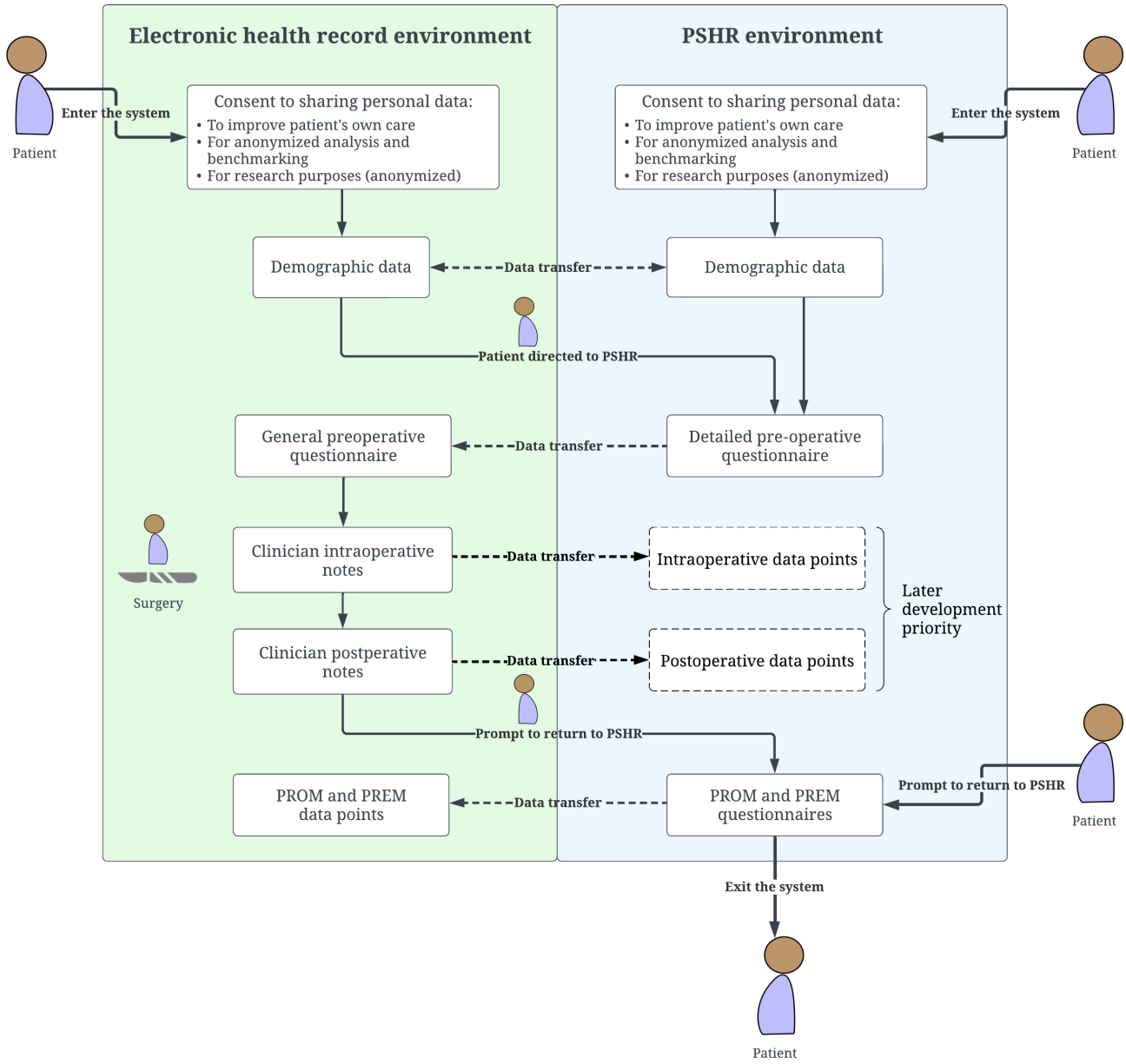
### Phase 2: Define

Insights from the initial interviews were that individuals presenting for surgical procedures have a need for connection with and feedback from their health care providers and a willingness to engage in actions necessary to navigate a challenging phase in their lives. Contrary to the investigators' expectations, participants who had used the PSHR were not discouraged by the length of the questionnaires, provided they perceived a clear and meaningful purpose to their completion. In addition, patients expressed a need for information related to their surgical procedures, highlighting the importance of incorporating targeted educational content to support informed decision-making throughout the perioperative journey.

### Phase 3: Ideate

The research team brainstormed suggestions and expectations from patient groups, and how the PSHR can be integrated into usual care processes. While standardized questionnaires remained unchanged, their sequence was reorganized to group similar questions, particularly in the PSHR preoperative questionnaire, where multiple risk assessments and surveys are consolidated into a single comprehensive questionnaire. Feedback messages were developed to provide patients with information tailored to their questionnaire responses. Various approaches were explored to support patient-clinician communication through the PSHR. The research team also considered the potential interaction of the PSHR with electronic health records (Figure 2).

**Figure 2.** Proposed PSHR interaction and interoperability with electronic health records. PREM: patient-reported experience measure; PROM: patient-reported outcomes measure; PSHR: Perioperative Shared Health Record.




## Phase 4: Prototype

A series of low-fidelity wireframes were created based on the potential solutions obtained during phase 3. These wireframes were refined into a low-fidelity paper prototype of the PSHR.

## Phase 5: User Test

Five individuals consented to participate in prototype testing. Three participants had previously used the PSHR ("expert users"), and 2 participants had no prior experience of the PSHR ("novice users"). Five users have been reported as sufficient for undertaking user testing [58,59]. Demographics are summarized in Table 3. All participants reported having their own mobile phone and usually accessing the internet and their email on their mobile phone and not on a computer. Each testing session took approximately 60 minutes to complete, with the most time spent on the second task. Figure 3 shows paper prototype testing in action; Figure 3A shows a participant discussing task 2 (completing the preoperative questionnaire), and Figure 3B shows a participant responding to a pop-up notification during task 4 (completing postoperative questionnaires).

**Table .** Demographic data for participants taking part in phase 5: prototype testing.

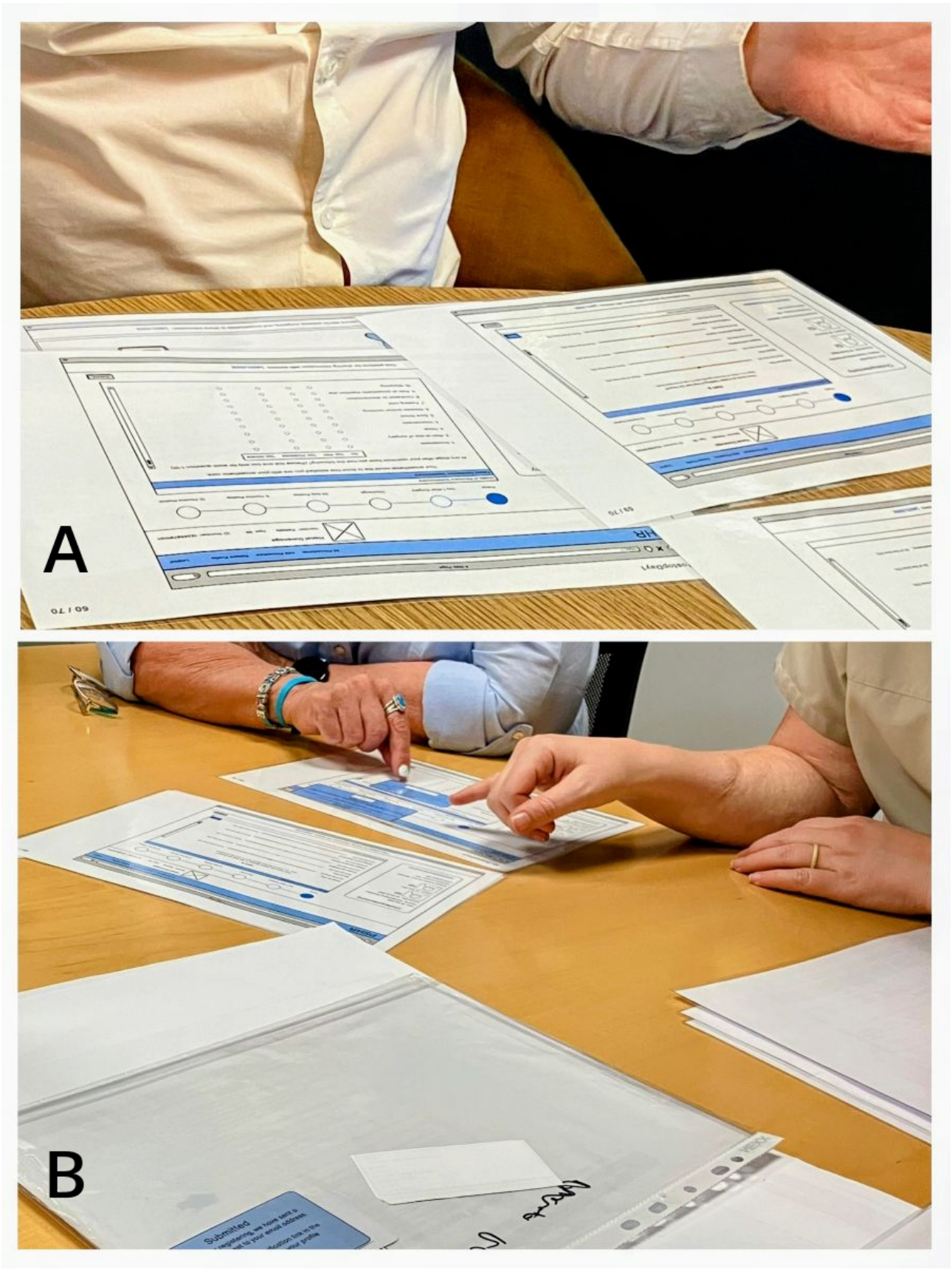
| ID[a] | Age (years) | Sex | Language | Race | Prior PSHR[b] experience | Education | Surgery |
|---|---|---|---|---|---|---|---|
| T1 | 42 | Female | English | White | Yes | After school qualification | Major abdominal |
| T2 | 63 | Female | Afrikaans | White | Yes | After school qualification | Major abdominal |
| T3 | 46 | Male | Afrikaans | White | Yes | After school qualification | N/A[c]—assisted family member |
| T4 | 21 | Male | Northern Sotho | Black | No | After school qualification | Head and neck |
| T5 | 32 | Male | Setswana | Black | No | Secondary school completed | Head and neck |

[a]User ID explanation: “T” denotes participant responses in the user testing phase.

[b]PSHR: Perioperative Shared Health Record.

[c]N/A: not applicable.

**Figure 3.** Paper prototype testing in action. (A) A participant discussing task 2 (completing the preoperative questionnaire). (B) A participant responding to a pop-up notification during task 4 (completing postoperative questionnaires).




For the first task (registering and consenting), trust was an important factor for 2 of the participants. Prior notification by their doctor to expect a registration email would help improve trust when receiving a link to an unknown website:

> *Yeah, knowing that this is safe, yes, because I spoke to you and I know that you will give me something like this. I think a personal call, direct, to say I'm sending you something now. I trust it more, because I know that I'm protected.* [T4]

Two of the participants read the consent form in detail, 2 participants scrolled through with minimal reading, and 1 participant indicated that he would abort the process when confronted with a lengthy consent process:

> *Unless if I'm buying a car, I don't read the details.* [T1]

> *As soon as I get this information, when I get to this one, I will say, yoh aha, this is too much. Pause, pause, pause.* [T4]

The 2 participants who read the consent form doubted that most users would engage deeply with the consent process. One participant suggested that using illustrations or icons could clarify abstract concepts. During the second task (preoperative questionnaire), various data input methods were tested, including radio buttons, colored numerical sliders, and free text blocks. Users seemed to appreciate color coding to interpret the numerical sliders. One participant suggested modifications to the order of questions.

Participants were able to navigate to the information portal (task 3); all participants found general information links useful but preferred procedure-specific content, with 3 favoring video links over text, and 1 mentioning that they would refer to written information only if the desired content was not available in the video links.

By the fourth task, participants were familiar with the layout of the home page and the questionnaires. Their understanding of the timeline had improved, but several suggestions were made to enhance its visual clarity. Feedback messages following questionnaire completion elicited mixed responses; some participants expressed concern that alerts about poor recovery outcomes could cause anxiety:

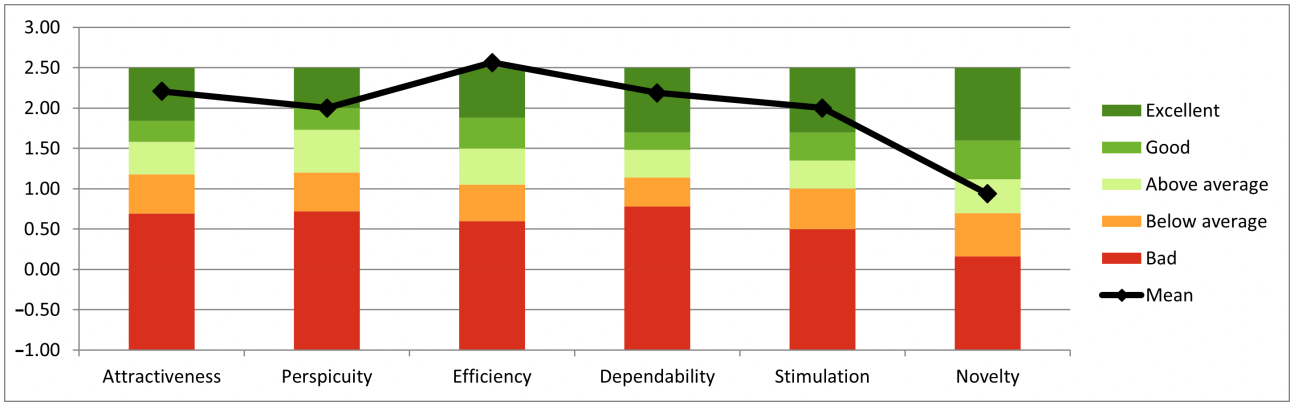
> *It makes me feel worried...I will go back to what I completed [to check] that I completed it correctly. Okay. Because there might be something that I said that might alarm the system.* [T5]

All participants indicated that they would value automatic feedback from their surgeon or anesthetist if they recorded poor scores on their postoperative questionnaires, with expectations for response times ranging from 30 to 60 minutes to up to 48 hours. Participants also noted that a lack of clinician feedback would reduce their motivation to continue using the platform. The potential for the PSHR to enhance patient-doctor relationships is illustrated in this quote:

> *So you still need to get to the buy-in from this. Where's my buy-in coming from? It's coming from the aftercare service, from the doctor, building that relationship. Because [of feedback through the PSHR] I've got a relationship again with the doctor, the surgeon or the rooms. I would be their patient for life! I think that's for me, that's how you get the buy-in to carry on the rest of the process.* [T2]

Four participants (3 experts and 1 novice) completed the SUS and UEQ questionnaires related to the paper prototype testing. The overall mean SUS score was 91.3 (SD 5.7), which indicates very good usability. The mean SUS scores per usability aspect were learnability 91.3 (SD 10.2), efficiency 93.4 (SD 6.9), and satisfaction 89.1 (SD 6.8). The UEQ mean scores for the attractiveness scale are 2.21 (SD 0.6), perspicuity 2.0 (SD 0.9), efficiency 2.6 (SD 0.4), dependability 2.2 (SD 0.6), stimulation 2.0 (SD 0.8), and novelty 0.94 (SD 1.4). The UEQ scores ranged from good to excellent, except for novelty, which scored above average (Figure 4).

**Figure 4.** Comparing the Perioperative Shared Health Record (PSHR) paper prototype User Experience Questionnaire scores to benchmark data. The measured scale means from our study are compared in relation to existing values from a benchmark dataset, which allows conclusions about the relative quality of the PSHR paper prototypes according to user testing compared with other products.




## Discussion

### Principal Results

This study applied a human-centered design approach to evaluate and improve patient user experience of a digital health tool developed to capture perioperative patient-reported outcomes. While identifying key usability challenges, it also showed how digital health tools such as the PSHR can help enhance connection between patients and clinicians through information sharing and timely feedback. The findings contribute new evidence from an LMIC setting, where practical integration of PROs into perioperative care is limited. By drawing on patient experiences as a resource for design, the study demonstrates how patient involvement can inform iterative

improvements to digital health tools and strengthen person-centered perioperative care [7,21,22,60,61].

The findings of our study align with established user experience principles in digital health design, emphasizing the importance of empathy, communication, accessibility, regulatory compliance, and data privacy and security [21].

Mapping the patient journey revealed the emotional strain of the perioperative period and the value of designing digital health tools that provide empathetic support during this vulnerable time [21,60]. This aligns with previous research highlighting that emotional engagement and good information provision are central to person-centered perioperative care and patient satisfaction [1,62,63]. Experiences reported in HIC show that access to targeted digital health tools can improve patient well-being and empowerment as well as improve postoperative outcomes [18,61,63,64]. Providing patients with a digital resource that offers clear, accessible information may therefore strengthen patient engagement by improving understanding of recovery and fostering a sense of partnership in care [1,18,65].

A key finding was the importance of communication and feedback from clinicians to create trust and to maintain motivation to continue using the system. This aligns with studies on implementation of electronic health records and patient perspectives on digital health tools [61,66]. It was interesting to note that patients were not deterred by lengthy questionnaires if they perceived them as purposeful. Participants valued knowing that their submitted data would inform their care, similar to evidence that perceived purpose and clinician responsiveness may increase adherence to digital platforms [61,64,67]. However, clinicians may not always see the value of using digital health tools to strengthen relationships with patients, especially if these tools are perceived as adding to their workload [28,66,68]. It is important to note that for patients who used the PSHR, a lack of clinician feedback following the completion of the postoperative questionnaires reduced their motivation to continue engaging with the tool. Therefore, it is important that patient needs are balanced with clinical feasibility. One potential solution would be to use automated alerts from PROM data that notify clinicians when patient responses are below a predefined threshold, prompting timely feedback to those patients who need them. This could enhance the perceived usefulness and reliability of the PSHR and strengthen the patient-clinician relationship, without overburdening the clinician [1].

Using the PSHR to track and benchmark recovery progress can offer reassurance or prompt patients to seek help when needed. Such features can promote self-management by helping patients to understand their recovery and to feel more in control of their health [61,64,67]. From our user testing, it emerged that when automated feedback messages to patients flag potential concerns, these messages should balance the communication of information with reassurance and clear guidance on next steps.

Despite the benefits of using a digital tool such as the PSHR, barriers such as limited digital literacy, the high cost of data, and inconsistent access to internet connectivity are significant obstacles to digital health implementation in South Africa [33,39] Patient preference for accessing the PSHR via WhatsApp links highlights the importance of incorporating widely used, low-barrier communication channels in LMIC settings. This aligns with priorities outlined in the South African Digital Health Strategy and the WHO Global Strategy on Digital Health, which highlights the need for equitable access, user-centered design, and interoperability of digital health tools [20,69].

Regulatory constraints, particularly around the consent process, present a design challenge for the PSHR. While simplifying consent forms with icons and condensed text may improve accessibility, maintaining careful attention to detail and to legal and ethical standards is important to ensure the integrity of the consent process. Providing reliable and up-to-date medical information is resource-intensive, whether creating original content or vetting existing material. One potential solution is to involve patients in content development and curation, fostering a collaborative platform. However, ensuring the accuracy of medical content would still require professional oversight and quality control.

An additional consideration in the design of the PSHR is safeguarding data privacy and security, especially as it collects personal and health information subject to the South African Protection of Personal Information Act, comparable with the Health Insurance Portability and Accountability Act in the United States and the General Data Protection Regulation in Europe. In the South African private health care sector, access to patient data can be restricted to the patient and their designated surgeon and anesthetist, which enhances data security. However, in the public health care sector, where care is provided by teams rather than individuals, maintaining data privacy may be more difficult. Concerns about cybersecurity and a lack of trust in an unknown system were seen by some patients as barriers to engaging with the PSHR.

### Strengths and Limitations

A key strength of this study lies in its adherence to design thinking and human-centered design principles [21,22]. The use of Karagianni's Optimized Honeycomb model provided a structured lens for analyzing user experience and expectations, capturing functional cognitive and emotional factors influencing user experience [47]. Involving individual patients in the design process enabled the research team to draw on their lived experience as a form of expertise. By prioritizing the needs of actual patients rather than relying on personas, we aim to advance our mission of developing an intuitive, efficient, user-friendly, and person-centered tool. Furthermore, the inclusion of a diverse sample of patients from both the public and private health care sectors in South Africa enhances the study's relevance, particularly as the country progresses toward universal health coverage. Including in the research team a clinician focused on patient-reported outcomes and a hospital patient representative kept the group focused on a patient-centered approach.

The main limitation of this project is the inclusion of a relatively small patient sample across the various phases, limiting the generalizability of findings. However, from a user experience research perspective, idea saturation in initial interviews suggests sufficient theme coverage, and prototype testing revealed consistent usability issues, aligning with Nielsen and

Landauer's 5-user rule [58,59]. The original study protocol aimed to include patients from Sweden to allow comparisons with a high-resource setting, but logistical and resource constraints prevented this. Future research will focus on expanding data collection and user testing, with a comparative analysis between HICs with well-established digital health platforms and LMICs where digital health systems are still being developed. An additional limitation was that user testing was conducted only with low-fidelity paper prototypes. However, this is a recognized approach within design thinking methodology, enabling iterative development without significant cost investment during the early stages when design elements remain subject to change [23,25,70].

Considering that one of the researchers has a professional interest in PROs and had established rapport with several participants, this may have introduced a subtle positive bias in how participants perceived and articulated the value of the PSHR. Furthermore, as the majority of patients had undergone intermediate or major surgical procedures, their emphasis on the need for information and emotional support may not be generalizable to patients presenting for minor operations.

Another limitation is that the study does not capture the user experience of health care providers. The original project plan included workshops with surgeons and anesthetists; however, time constraints necessitated postponing these activities. Ongoing work by the South African authors includes intentional network weaving to promote data-driven surgery which will include engagement with perioperative clinicians.

Although initially unfamiliar to the anesthesiologist investigators, the qualitative and user experience methodologies provided valuable learning. This collaboration highlighted the importance of such approaches in helping clinicians understand patient needs and to develop intuitive digital health tools.

### Future Research

Findings from this study will inform further design iterations of the PSHR, both to optimize its use in individual patient care and to generate future research outputs. Next steps include testing a high-fidelity prototype and evaluating the final product in real care settings, with particular attention to patient experiences over time: for example, how patients respond to repeated questionnaires that may appear similar at different intervals. Data from the PSHR will in time become a resource for organizational development and quality improvement. Ongoing development will require the active involvement of both clinicians and patients to ensure that the tool remains relevant, feasible, and responsive to real-world clinical processes and workflows.

### Conclusions

This study is one of the first to apply human-centered design principles to a perioperative digital health tool in an LMIC setting, addressing usability challenges and patient engagement. Key user experience factors influencing patient engagement included communication, feedback, and access to information throughout the surgical journey. Digital health tools such as the PSHR can strengthen communication and support person-centered perioperative care by integrating PROs into clinical workflows and care processes. As health care systems worldwide move toward digital integration, our findings provide valuable insights into the factors to consider when digital health tools are introduced in diverse health care contexts. Future research should focus on integrating digital health tools into clinical workflows and assessing their impact on person-centered outcomes and care delivery, with particular emphasis on involving all relevant stakeholders, both clinicians and patients, to ensure that the tools are contextually appropriate and aligned with real-world processes and workflow needs.

## Funding

This work is based on the research supported in part by the National Research Foundation of South Africa (reference no. SASUF23041894320) through a seed grant for collaborative research with the South African Swedish University Forum (SASUF-NRF). The research reported in this manuscript is partially supported by the South African Medical Research Council (SAMRC) through its Division of Research Capacity Development under the Research Capacity Development Initiative from funding received from the South African National Treasury. The content and findings reported/illustrated are the sole deduction, view, and responsibility of the researcher, and do not reflect the official position and sentiments of the SAMRC/SASUF-NRF.

## Authors' Contributions

CS participated in study concept and design, data collection, data analysis, data interpretation, writing of the manuscript, and critical revision of manuscript. HD contributed to data collection, data analysis, data interpretation, and critical revision of manuscript. CJO and GF participated in study concept and design, data collection, data analysis, data interpretation, and critical revision of manuscript. MC and HK participated in study concept and design and critical revision of manuscript.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Patient semistructured interviews guide.
[DOCX File, 17 KB - periop_v9i1e79349_app1.docx ]

Multimedia Appendix 2
Key tasks during paper prototype testing.
[DOCX File, 21 KB - periop_v9i1e79349_app2.docx ]

Multimedia Appendix 3
Detailed thematic analysis.
[DOCX File, 57 KB - periop_v9i1e79349_app3.docx ]

## References


1. Bavin E, Tobiano G, Gillespie BM. Patient experience of elective general surgery: an integrative review. Collegian 2023 Oct;30(5):676-685. [doi: 10.1016/j.colegn.2023.08.002]
2. Wang XS, Gottumukkala V. Patient-reported outcomes: Is this the missing link in patient-centered perioperative care? Best Pract Res Clin Anaesthesiol 2021 Dec;35(4):565-573. [doi: 10.1016/j.bpa.2020.10.006] [Medline: 34801218]
3. Steenkamp L, Silva SA, Collie S. 8-year review of patient satisfaction surveys: Medical scheme members who score initial hospital stay poorly, more likely to have unplanned readmission. 2024. URL: https://www.discovery.co.za/corporate/health-insights-review-patient-surveys-hospitals [accessed 2024-05-02]
4. Roberts GP, Levy N, Lobo DN. Patient-centric goal-oriented perioperative care. Br J Anaesth 2021 Mar;126(3):559-564. [doi: 10.1016/j.bja.2020.12.004] [Medline: 33419527]
5. Myles PS, Boney O, Botti M, et al. Systematic review and consensus definitions for the Standardised Endpoints in Perioperative Medicine (StEP) initiative: patient comfort. Br J Anaesth 2018 Apr;120(4):705-711. [doi: 10.1016/j.bja.2017.12.037] [Medline: 29576111]
6. Billig JI, Sears ED, Travis BN, Waljee JF. Patient-reported outcomes: understanding surgical efficacy and quality from the patient's perspective. Ann Surg Oncol 2020 Jan;27(1):56-64. [doi: 10.1245/s10434-019-07748-3] [Medline: 31489556]
7. Ekman I, Swedberg K, Taft C, et al. Person-centered care--ready for prime time. Eur J Cardiovasc Nurs 2011 Dec;10(4):248-251. [doi: 10.1016/j.ejcnurse.2011.06.008] [Medline: 21764386]
8. Leplege A, Gzil F, Cammelli M, Lefeve C, Pachoud B, Ville I. Person-centredness: conceptual and historical perspectives. Disabil Rehabil 2007;29(20-21):1555-1565. [doi: 10.1080/09638280701618661] [Medline: 17922326]
9. Knutsson B, Kadum B, Eneqvist T, Mukka S, Sayed-Noor AS. Patient satisfaction with care is associated with better outcomes in function and pain 1 year after lumbar spine surgery. J Patient Cent Res Rev 2022;9(1):7-14. [doi: 10.17294/2330-0698.1883] [Medline: 35111878]
10. Çakmak C, Uğurluoğlu Ö. The effects of patient-centered communication on patient engagement, health-related quality of life, service quality perception and patient satisfaction in patients with cancer: a cross-sectional study in Türkiye. Cancer Control 2024;31:10732748241236327. [doi: 10.1177/10732748241236327] [Medline: 38411086]
11. Olsson LE, Karlsson J, Ekman I. The integrated care pathway reduced the number of hospital days by half: a prospective comparative study of patients with acute hip fracture. J Orthop Surg Res 2006 Sep 25;1(1):3. [doi: 10.1186/1749-799X-1-3] [Medline: 17150123]
12. Kingsley C, Patel S. Patient-reported outcome measures and patient-reported experience measures. BJA Educ 2017 Apr;17(4):137-144. [doi: 10.1093/bjaed/mkw060]
13. Weldring T, Smith SMS. Patient-reported outcomes (PROs) and patient-reported outcome measures (PROMs). Health Serv Insights 2013;6(6):61-68. [doi: 10.4137/HSI.S11093] [Medline: 25114561]
14. Moonesinghe SR, Jackson AIR, Boney O, et al. Systematic review and consensus definitions for the Standardised Endpoints in Perioperative Medicine initiative: patient-centred outcomes. Br J Anaesth 2019 Nov;123(5):664-670. [doi: 10.1016/j.bja.2019.07.020] [Medline: 31493848]
15. Myles PS, Grocott MPW, Boney O, Moonesinghe SR, COMPAC-StEP Group. Standardizing end points in perioperative trials: towards a core and extended outcome set. Br J Anaesth 2016 May;116(5):586-589. [doi: 10.1093/bja/aew066] [Medline: 27106961]
16. Sokas C, Hu F, Edelen M, Sisodia R, Pusic A, Cooper Z. A review of PROM implementation in surgical practice. Ann Surg 2022 Jan 1;275(1):85-90. [doi: 10.1097/SLA.0000000000005029] [Medline: 34183512]
17. Tsang C, Lee KS, Richards H, Blazeby JM, Avery KNL. Electronic collection of patient-reported outcomes following discharge after surgery: systematic review. BJS Open 2021 Mar 5;5(2):zraa072. [doi: 10.1093/bjsopen/zraa072] [Medline: 33782708]
18. van der Meij E, Anema JR, Otten RHJ, Huirne JAF, Schaafsma FG. The effect of perioperative E-health interventions on the postoperative course: a systematic review of randomised and non-randomised controlled trials. PLoS One 2016;11(7):e0158612. [doi: 10.1371/journal.pone.0158612] [Medline: 27383239]
19. Coulter A, Locock L, Ziebland S, Calabrese J. Collecting data on patient experience is not enough: they must be used to improve care. BMJ 2014 Mar 26;348:g2225. [doi: 10.1136/bmj.g2225] [Medline: 24671966]
20. Global strategy on digital health 2020-2025. World Health Organization. 2021. URL: https://iris.who.int/server/api/core/bitstreams/1f4d4a08-b20d-4c36-9148-a59429ac3477/content [accessed 2024-08-11]

21. Familoni BT, Odetunde Babatunde S. User experience (UX) design in medical products: theoretical foundations and development best practices. Eng Sci technol j 2024;5(3):1125-1148. [doi: 10.51594/estj.v5i3.975]
22. Melles M, Albayrak A, Goossens R. Innovating health care: key characteristics of human-centered design. Int J Qual Health Care 2021 Jan 12;33(Suppl 1):37-44. [doi: 10.1093/intqhc/mzaa127] [Medline: 33068104]
23. Krolikowski KA, Bi M, Baggott CM, Khorzad R, Holl JL, Kruser JM. Design thinking to improve healthcare delivery in the intensive care unit: promise, pitfalls, and lessons learned. J Crit Care 2022 Jun;69:153999. [doi: 10.1016/j.jcrc.2022.153999] [Medline: 35104692]
24. Doorley S, Holcomb S, Klebahn P, Segovia K, Utley J. Design thinking bootleg. In: Elbogen E, editor. Standford Design School 2018. URL: https://dschool.sfo3.digitaloceanspaces.com/documents/dschool_bootleg_deck_2018_final_sm2-6.pdf [accessed 2025-12-12]
25. An introduction to design thinking: process guide. Institute of Design at Stanford. URL: https://web.stanford.edu/~mshanks/MichaelShanks/files/509554.pdf [accessed 2025-03-02]
26. Gilbert RM. Reimagining digital healthcare with a patient-centric approach: the role of user experience (UX) research. Front Digit Health 2022;4(August):899976. [doi: 10.3389/fdgth.2022.899976] [Medline: 36016600]
27. Essén A, Scandurra I, Gerrits R, et al. Patient access to electronic health records: differences across ten countries. Health Policy Technol 2018 Mar;7(1):44-56. [doi: 10.1016/j.hlpt.2017.11.003]
28. Cijvat CD, Cornet R, Hägglund M. Factors influencing development and implementation of patients' access to electronic health records-a comparative study of Sweden and the Netherlands. Front Public Health 2021;9:621210. [doi: 10.3389/fpubh.2021.621210] [Medline: 34169054]
29. Wiljer D, Urowitz S, Apatu E, et al. Patient accessible electronic health records: exploring recommendations for successful implementation strategies. J Med Internet Res 2008 Oct 31;10(4):e34. [doi: 10.2196/jmir.1061] [Medline: 18974036]
30. Hägglund M, Scandurra I. Patients' online access to electronic health records: current status and experiences from the implementation in Sweden. Stud Health Technol Inform 2017;245:723-727. [Medline: 29295193]
31. Emilsson L, Lindahl B, Koster M, Lambe M, Ludvigsson JF. Review of 103 Swedish healthcare quality registers. Int J Epidemiol 2015 Oct;44(suppl_1):i174-i174. [doi: 10.1093/ije/dyv096.249]
32. Bridgwood A. General household survey. J Public Health Med 1999 Dec;21(4):482. [Medline: 11469378]
33. Khumalo MH, Moloi TS. Barriers to digital transformation in Gauteng's municipal health clinics. J Local Gov Res Innov 2025 Aug 22;6:1-12. [doi: 10.4102/jolgri.v6i0.282]
34. Reddy CL, Makasa EM, Biccard B, et al. Surgery as a component of universal healthcare: where is South Africa? S Afr Med J 2019 Aug 28;109(9):624-625. [doi: 10.7196/SAMJ.2019.v109i9.14233] [Medline: 31635583]
35. Coovadia H, Jewkes R, Barron P, Sanders D, McIntyre D. The health and health system of South Africa: historical roots of current public health challenges. Lancet 2009 Sep 5;374(9692):817-834. [doi: 10.1016/S0140-6736(09)60951-X] [Medline: 19709728]
36. Day C, Gray A, Cois A, Ndlovu N, Massyn N, Boerma T. Is South Africa closing the health gaps between districts? Monitoring progress towards universal health service coverage with routine facility data. BMC Health Serv Res 2021 Sep 13;21(Suppl 1):194. [doi: 10.1186/s12913-021-06171-3] [Medline: 34511085]
37. Mayosi BM, Lawn JE, van Niekerk A, et al. Health in South Africa: changes and challenges since 2009. Lancet 2012 Dec 8;380(9858):2029-2043. [doi: 10.1016/S0140-6736(12)61814-5] [Medline: 23201214]
38. Thabethe SS, Mathews VE. Quality of routine health data in DHIS2 in South Africa: Eastern Cape province from 2017–2020. S Afr J Inf Manag 2024;26(1):1903. [doi: 10.4102/sajim.v26i1.1903]
39. Zharima C, Griffiths F, Goudge J. Exploring the barriers and facilitators to implementing electronic health records in a middle-income country: a qualitative study from South Africa. Front Digit Health 2023;5:1207602. [doi: 10.3389/fdgth.2023.1207602] [Medline: 37600481]
40. Kluyts HL. Strengthening surgical systems in LMICs: data-driven approaches. Lancet Global Health 2024 Nov;12(11):e1744-e1745. [doi: 10.1016/S2214-109X(24)00375-9] [Medline: 39245052]
41. Nicol E, Hanmer LA, Mukumbang FC, Basera W, Zitho A, Bradshaw D. Is the routine health information system ready to support the planned national health insurance scheme in South Africa? Health Policy Plan 2021 Jun 1;36(5):639-650. [doi: 10.1093/heapol/czab008] [Medline: 33822055]
42. Matshidze P, Hanmer L. Health information systems in the private health sector. In: South African Health Review: Health Systems Trust; 2007. URL: https://www.hst.org.za/publications/South%20African%20Health%20Reviews/11_SAHR_2006-2007_Section%206.pdf [accessed 2025-12-12]
43. Transforming surgical care in Africa through data-driven innovation. Safe Surgery SA. URL: https://safesurgery.co.za [accessed 2020-10-15]
44. Perioperative shared health record. Safe Surgery SA. URL: https://safesurgery.co.za/pshr/ [accessed 2023-04-08]
45. 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 Apr;42(2):377-381. [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
46. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]

47. Karagianni K. Optimizing the UX honeycomb. Medium. 2018. URL: https://uxdesign.cc/optimizing-the-ux-honeycomb-1d10cfb38097 [accessed 2024-10-13]
48. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
49. Im D, Pyo J, Lee H, Jung H, Ock M. Qualitative research in healthcare: data analysis. J Prev Med Public Health 2023 Mar;56(2):100-110. [doi: 10.3961/jpmph.22.471] [Medline: 37055353]
50. Mansson L, Wiklund M, Öhberg F, Danielsson K, Sandlund M. Co-creation with older adults to improve user-experience of a smartphone self-test application to assess balance function. Int J Environ Res Public Health 2020 May 26;17(11):3768. [doi: 10.3390/ijerph17113768] [Medline: 32466484]
51. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today 2004 Feb;24(2):105-112. [doi: 10.1016/j.nedt.2003.10.001] [Medline: 14769454]
52. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the System Usability Scale. Int J Hum Comput Interact 2008 Jul 29;24(6):574-594. [doi: 10.1080/10447310802205776]
53. Brooke J. SUS: A “quick and dirty” usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. Usability Evaluation in Industry, 1st edition: Taylor & Francis; 1996.
54. Hyzy M, Bond R, Mulvenna M, et al. System Usability Scale benchmarking for digital health apps: meta-analysis. JMIR Mhealth Uhealth 2022 Aug 18;10(8):e37290. [doi: 10.2196/37290] [Medline: 35980732]
55. Laugwitz B, Held T, Schrepp M. Construction and Evaluation of a User Experience Questionnaire: Springer; 2008, Vol. 5298 LNCS:63-76. [doi: 10.1007/978-3-540-89350-9_6]
56. A fast and reliable questionnaire to measure the User Experience of interactive products. User Experience Questionnaire (UEQ). URL: https://www.ueq-online.org/ [accessed 2025-01-19]
57. Schrepp M, Hinderks A, Thomaschewski J. Construction of a benchmark for the User Experience Questionnaire (UEQ). Int J Interact Multim Artif Intell 2017;4(4):40. [doi: 10.9781/ijimai.2017.445]
58. Nielsen J, Landauer TK. A mathematical model of the finding of usability problems. 1993 Presented at: Conference on Human Factors in Computing Systems—Proceedings Published by ACM; Apr 24-29, 1993 URL: http://portal.acm.org/citation.cfm?doid=169059 [accessed 2025-12-19] [doi: 10.1145/169059.169166]
59. Nielsen J. Why you only need to test with 5 users. Nielsen Norman Group. 2000. URL: https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/ [accessed 2025-01-19]
60. Molina-Recio G, Molina-Luque R, Jiménez-García AM, Ventura-Puertos PE, Hernández-Reyes A, Romero-Saldaña M. Proposal for the user-centered design approach for health apps based on successful experiences: integrative review. JMIR Mhealth Uhealth 2020 Apr 22;8(4):e14376. [doi: 10.2196/14376] [Medline: 32319965]
61. Madanian S, Nakarada-Kordic I, Reay S, Chetty T. Patients’ perspectives on digital health tools. PEC Innov 2023;2:100171. [doi: 10.1016/j.pecinn.2023.100171]
62. Jones CH, O’Neill S, McLean KA, Wigmore SJ, Harrison EM. Patient experience and overall satisfaction after emergency abdominal surgery. BMC Surg 2017;17(1):76. [doi: 10.1186/s12893-017-0271-5]
63. Sansalone E, White J, Zucca A, et al. Acceptability of the RecoverEsupport digital health intervention among patients undergoing breast cancer surgery: qualitative study. J Med Internet Res 2025 Oct 17;27:e77567. [doi: 10.2196/77567] [Medline: 41105868]
64. ElKefi S, Asan O. How technology impacts communication between cancer patients and their health care providers: a systematic literature review. Int J Med Inform 2021 May;149:104430. [doi: 10.1016/j.ijmedinf.2021.104430] [Medline: 33684711]
65. Hladkowicz E, Yachnin D, Boland L, et al. Evaluation of a preoperative personalized risk communication tool: a prospective before-and-after study. Can J Anaesth 2020 Dec;67(12):1749-1760. [doi: 10.1007/s12630-020-01809-y] [Medline: 32929659]
66. Hägglund M, Scandurra I. Usability of the Swedish accessible electronic health record: qualitative survey study. JMIR Hum Factors 2022 Jun 23;9(2):e37192. [doi: 10.2196/37192] [Medline: 35737444]
67. Esper SA, Holder-Murray J, Meister KA, et al. A novel digital health platform with health coaches to optimize surgical patients: feasibility study at a large academic health system. JMIR Perioper Med 2024 Apr 4;7(1):e52125. [doi: 10.2196/52125] [Medline: 38573737]
68. Ahmed A, Ho CW, Grant Y, Archer S, Carrington EV. Acceptability of digital health interventions in perioperative care: a systematic review and narrative synthesis of clinician perspectives. BMJ Open 2025 Mar 15;15(3):e086412. [doi: 10.1136/bmjopen-2024-086412] [Medline: 40090692]
69. National digital health strategy for South Africa 2019-2024. National Department of Health. 2019 May. URL: https://knowledgehub.health.gov.za/system/files/elibdownloads/2023-04/national-digital-strategy-for-south-africa-2019-2024-b.pdf [accessed 2025-10-20]
70. Altman M, Huang TTK, Breland JY. Design thinking in health care. Prev Chronic Dis 2018 Sep 27;15:E117. [doi: 10.5888/pcd15.180128] [Medline: 30264690]

## Abbreviations

**HIC:** high-income country
**LMIC:** low- and middle-income country
**PREM:** patient-reported experience measure
**PRO:** patient-reported outcome
**PROM:** patient-reported outcome measure
**PSHR:** Perioperative Shared Health Record
**REDCap:** Research Electronic Data Capture
**SACSOS:** South African Collaborative Surgical Outcomes Study
**SUS:** System Usability Scale
**UEQ:** User Experience Questionnaire

*Edited by J Shiffermiller; submitted 19.Jun.2025; peer-reviewed by J Ohde, S Nilsson; revised version received 28.Oct.2025; accepted 18.Nov.2025; published 12.Jan.2026.*

*Please cite as:*
*Steyl C, Orre C, Foster G, Duvenage H, Chew MS, Kluyts HL*
*Enhancing the User Experience of a Perioperative Digital Health Tool for Information Exchange Using a Human-Centered Design Thinking Approach: Qualitative Observational Study*
*JMIR Perioper Med 2026;9:e79349*
*URL: https://periop.jmir.org/2026/1/e79349*
*doi:10.2196/79349*

# Clinical Feasibility and Outcomes of Surgeon-Performed Laparoscopic-Guided Subcostal Transversus Abdominis Plane Block in Laparoscopic Cholecystectomy: Prospective Observational Study

Sarun Mahasupachai, MD; Thawatchai Tullavardhana, MD

Department of Surgery, Faculty of Medicine, Srinakharinwirot University, Ransit - Nakhon nayok Road, Ongkharak, Nakhon nayok, Thailand

**Corresponding Author:**
Thawatchai Tullavardhana, MD
Department of Surgery, Faculty of Medicine, Srinakharinwirot University, Ransit - Nakhon nayok Road, Ongkharak, Nakhon nayok, Thailand

## Abstract

**Background:** Laparoscopic-guided subcostal transversus abdominis plane (TAP) block has been introduced as a surgeon-performed approach to postoperative analgesia in laparoscopic cholecystectomy (LC), allowing direct visual confirmation of local anesthetic delivery without ultrasound guidance. However, evidence regarding its clinical outcomes, particularly in patients with complicated gallstone disease, remains limited.

**Objective:** This study aimed to evaluate postoperative analgesic outcomes and identify factors associated with opioid requirement following laparoscopic-guided subcostal TAP block.

**Methods:** A prospective observational study was conducted between November 2023 and October 2024 at Srinakharinwirot University Hospital, Thailand. Patients (aged 18 - 80 years) undergoing LC for uncomplicated or complicated gallstone disease received a laparoscopic-guided subcostal TAP block with 0.25% bupivacaine. Postoperative pain was assessed using the Visual Analog Scale at 2, 4, 6, 8, 12, and 24 hours. Morphine administration within the first 24 hours was recorded. Associations between perioperative variables and opioid requirement were analyzed using univariate and exploratory multivariable logistic regression.

**Results:** A total of 42 patients were included in the analysis. Of these, 21 (50%) did not require postoperative opioids, while the remaining patients (n=21, 50%) received a mean cumulative morphine dose of 3.86 (SD 1.39) mg within 24 hours. Pain scores were lower during the early postoperative period (2, 4, and 12 h) in patients who did not require opioids. Higher American Society of Anesthesiologists classification was independently associated with postoperative morphine requirement (odds ratio 6.51, 95% CI 1.37 - 30.96; *P*=.01). No major complications or local anesthetic toxicity were observed.

**Conclusions:** In this prospective observational cohort, laparoscopic-guided subcostal TAP block may be associated with favorable early postoperative analgesic profiles and relatively low opioid requirements after LC, including in patients with gallstone-related complications. Higher American Society of Anesthesiologists classification may be associated with increased opioid demand, highlighting the importance of individualized, risk-adapted analgesic strategies. Although limited by the absence of a control group and modest sample size, these findings support the clinical feasibility of surgeon-performed TAP block for consideration within multimodal analgesia approaches in enhanced recovery after surgery–oriented perioperative care.

**Trial Registration:** Thai Clinical Trials Registry TCTR20250314002; https://www.thaiclinicaltrials.org/show/TCTR20250314002

***(JMIR Perioper Med 2026;9:e87622)*** doi:10.2196/87622

**KEYWORDS**

laparoscopic cholecystectomy; minimally invasive surgery; postoperative pain; transversus abdominis plane block; enhanced recovery after surgery

## Introduction

Laparoscopic cholecystectomy (LC) is the standard surgical approach for gallstone disease, offering distinct advantages over open cholecystectomy, including shorter recovery times and earlier return to normal activities [1]. Despite these benefits, LC is associated with moderate to severe postoperative pain, particularly within the first 24 hours, often necessitating opioid analgesia [2]. High-dose opioid use, however, is frequently complicated by nausea, vomiting, dizziness, abdominal distension, and urinary retention, which may delay recovery and prolong hospitalization [3,4].

Enhanced recovery after surgery (ERAS) protocols have been widely implemented to optimize perioperative care and expedite

recovery. Multimodal analgesia represents a cornerstone of these protocols, aiming to minimize opioid use while maintaining effective pain control [5,6]. Within this framework, the subcostal transversus abdominis plane (TAP) block has emerged as a valuable component of multimodal analgesia, providing targeted pain relief following LC. When performed under ultrasound guidance using 0.25% bupivacaine, this block reliably anesthetizes thoracic (T7-T12) and lumbar (L1) nerves, thereby improving pain control and facilitating earlier mobilization [7,8]. Nevertheless, its dependence on anesthesiologist expertise and specialized equipment limits feasibility in certain clinical environments.

To address these limitations, the laparoscopic-guided subcostal TAP block has been developed as a surgeon-performed technique seamlessly incorporated into the operative workflow. Under direct laparoscopic visualization, local anesthetic can be precisely delivered into the TAP, providing consistent parietal analgesia while obviating the need for ultrasound equipment or additional personnel. This method has been demonstrated to be safe, efficient, and time-effective, offering a practical alternative for postoperative pain control following LC [9].

Nevertheless, most previous studies evaluating TAP block for LC have primarily focused on patients with uncomplicated gallstone disease and were conducted in controlled trial settings. A recent systematic review and meta-analysis demonstrated that TAP block is effective in reducing postoperative pain and opioid consumption after LC, with most included studies using ultrasound-guided techniques [10]. In contrast, this study evaluates a surgeon-performed, laparoscopic-guided subcostal TAP block integrated into routine operative workflow.

This study aimed to assess the clinical feasibility and outcomes of surgeon-performed, laparoscopic-guided subcostal TAP block for postoperative pain management in patients undergoing LC for both uncomplicated and complicated gallstone disease, including acute cholecystitis and biliary tract obstruction. Additionally, perioperative predictors of postoperative opioid requirement were explored.

## Methods

### Overview

A single-center observational study was conducted at the Department of Surgery, Faculty of Medicine, Srinakharinwirot University, Thailand, between November 2023 and October 2024. Eligible patients were aged 18 to 80 years and diagnosed with symptomatic cholelithiasis or gallstone-related complications. Uncomplicated gallstone disease was defined as symptomatic cholelithiasis or chronic cholecystitis without evidence of systemic inflammation or biliary complications. Complicated gallstone disease was defined as gallstone-related conditions associated with acute inflammation or biliary obstruction, including acute cholecystitis, acute cholangitis, or biliary obstruction requiring endoscopic retrograde cholangiopancreatography. Patients who were converted to open surgery or had a known allergy to bupivacaine were excluded. The target sample size was 40 patients; however, to enhance reliability, a total of 50 patients were enrolled.

### Study Design Declaration

The initial ethics-approved protocol was designed as a randomized comparison between subcostal TAP block and port-site local infiltration. However, due to limited patient recruitment, randomization could not be executed. This report therefore represents an observational analysis of patients who received the surgeon-performed, laparoscopic-guided subcostal TAP block in accordance with the originally approved protocol. No additional procedures, interventions, or deviations from the ethics approval were undertaken.

### Ethical Considerations

This study was approved by the Institutional Ethics Committee of Srinakharinwirot University (ethics code: SWUEC-004/2566F). Written informed consent was obtained from all participants prior to enrollment, and all patient data were collected and managed in accordance with institutional and international standards for data confidentiality and ethical research practice.

The study was retrospectively registered with the Thai Clinical Trials Registry (TCTR20250314002) following a change in study execution from the originally approved randomized protocol to a prospective observational design. Importantly, no protocol deviations occurred beyond the scope approved by the institutional ethics committee.

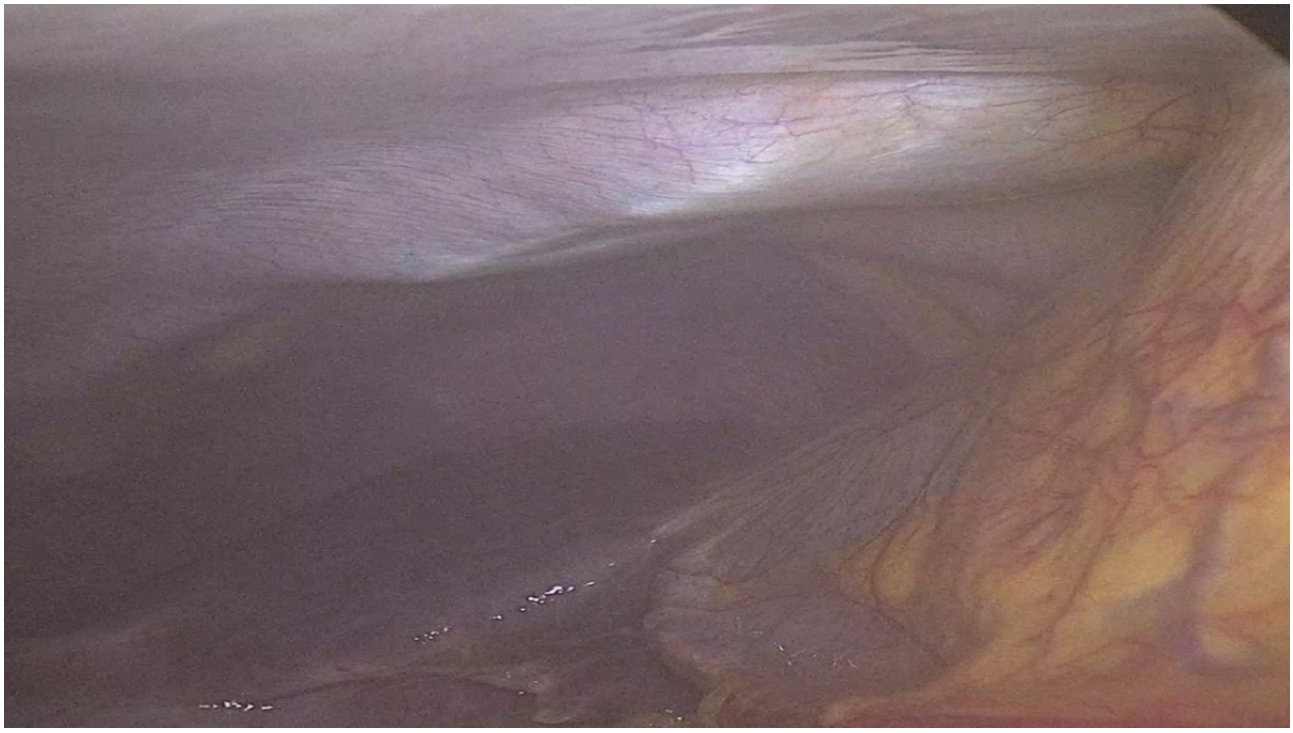
### Laparoscopic-Guided Subcostal TAP Block Technique

LC was performed under general anesthesia. Intraoperative analgesia was standardized according to the institutional protocol and was not analyzed separately. Prior to skin incision, 5 mL of 0.25% bupivacaine was infiltrated at the umbilical site for local analgesia. A 12-mm umbilical port was inserted for the laparoscopic camera, and pneumoperitoneum was established at an intra-abdominal pressure of 8 to 12 mmHg.

After establishment of pneumoperitoneum, the right upper quadrant was inspected under laparoscopic visualization. The injection point was identified at the right subcostal region, approximately 2 to 3 cm inferior to the costal margin and lateral to the midline, corresponding to the junction between the posterior rectus sheath and the transversus abdominis muscle.

Under direct laparoscopic visualization, a long spinal needle was inserted percutaneously toward the TAP, with initial advancement at an angle of approximately 60 to 80 degrees relative to the abdominal wall. Upon reaching the target fascial plane, the needle angle was adjusted to approximately 30 degrees to facilitate controlled anesthetic delivery. After negative aspiration, 20 mL of 0.25% bupivacaine was injected incrementally. Correct local anesthetic deposition was inferred from laparoscopic visualization of the Doyle bulge, indicating separation of the posterior rectus sheath and transversus abdominis muscle as a surrogate marker of appropriate fascial plane injection (Figure 1).

**Figure 1.** Illustration of Doyle bulge observed during laparoscopic-guided subcostal transversus abdominis plane block.



In our contemporary LC practice, routine use of a 12-mm epigastric working port has been replaced by three 5-mm working ports positioned along the right subcostal margin, in addition to the 12-mm umbilical optical port. This port configuration reflects an evolving minimally invasive approach aimed at reducing port-site trauma while maintaining adequate operative ergonomics and visualization. After cholecystectomy, the gallbladder was retrieved in a sterile bag through the 12-mm umbilical port. A supplementary video demonstrating the laparoscopic-guided subcostal TAP block technique is provided in Multimedia Appendix 1.

## Postoperative Pain Management and Monitoring

All patients received standardized postoperative analgesia. Oral acetaminophen (500 mg every 6 hours) and naproxen (250 mg twice daily) were prescribed as first-line agents; tramadol (50 mg twice daily) was substituted for patients with nonsteroidal anti-inflammatory drug intolerance. Persistent pain with a Visual Analog Scale (VAS) score >5 despite oral medication was managed with intravenous morphine (0.1 mg/kg every 4 hours as needed). Ondansetron (4 mg every 8 hours as required) was given for nausea and vomiting prophylaxis.

Pain intensity was evaluated using the VAS (0=no pain and 10=worst pain) at 2, 4, 6, 8, 12, and 24 hours postoperatively and during mobilization. Total morphine consumption and opioid-related adverse events (nausea, vomiting, and urinary retention) were recorded within the first 24 hours. Patients were observed for possible bupivacaine toxicity during this period.

## Clinical Variables

Data collection included patient demographics, comorbidities, laboratory parameters, operative details, postoperative outcomes, and histopathological findings. Laboratory parameters analyzed in this study were preoperative values obtained within 24 hours before surgery. Operative variables comprised surgical duration, estimated blood loss, and intraoperative complications.

Postoperative variables included pain scores, total morphine consumption, and postoperative complications. All data were prospectively recorded in a predesigned spreadsheet (Microsoft Excel) for statistical analysis. The dataset was stored in password-protected files on secure institutional computers, with access restricted to study investigators only. Patient identifiers were removed and replaced with coded study numbers to ensure confidentiality.

## Statistical Analysis

Statistical analysis was conducted using SPSS Statistics (version 27.0; IBM Corp). Descriptive statistics (mean, SD, frequencies, and percentages) were used to summarize demographic and clinical data. Comparisons between subgroups were conducted using the chi-square or Fisher exact test for categorical variables and the independent 2-tailed *t* test (or 1-way ANOVA where applicable) for continuous variables.

Exploratory multivariable logistic regression was applied to assess potential associations between perioperative factors and postoperative morphine requirement, acknowledging the limited sample size. Variables included in the multivariable model were selected based on clinical relevance and prior literature, rather than solely on statistical significance in univariate analyses. Given the limited number of outcome events, the analysis was conducted with a restricted number of covariates to maintain an acceptable events per variable ratio and minimize the risk of overfitting. A 2-tailed *P* value <.05 was considered statistically significant.

# Results

## Overview

During the study period, 50 patients underwent LC. In total, 8 (16%) patients were excluded due to conversion to open surgery, leaving 42 (84%) patients for analysis. Of these, 21 (50%) patients required postoperative morphine within the first 24 hours. Among patients who received opioids (n=21, 50%), the mean cumulative morphine dose was 3.86 (SD 1.39) mg.

## Patient Characteristics

Baseline demographic and clinical characteristics are summarized in Table 1. There were no significant differences between the morphine-required and morphine-free groups in age (*P*=.55), BMI (*P*=.55), or sex distribution (*P*=.22). Comorbid conditions were more frequent in the morphine-required group; however, this difference did not reach statistical significance (*P*=.10). The mean American Society of Anesthesiologists (ASA) classification was significantly higher among patients who required morphine (mean 2.14, SD 0.57 vs mean 1.67, SD 0.73; *P*=.024), reflecting a higher baseline perioperative risk profile.

**Table .** Baseline demographic and clinical characteristics of the study population.

| Variable | Morphine required (n=21) | Morphine free (n=21) | *P* value |
|---|---|---|---|
| Age (years), mean (SD) | 57.1 (14.1) | 54.3 (15.8) | .55 |
| Sex (female), n (%) | 13 (61.9) | 8 (38.1) | .22 |
| BMI (kg/m²), mean (SD) | 25.8 (5.3) | 24.8 (4.9) | .55 |
| ASA[a] classification, mean (SD) | 2.14 (0.57) | 1.67 (0.73) | .02 |
| Any comorbidity, n (%) | 17 (81.0) | 10 (47.6) | .10 |
| Diabetes mellitus, n (%) | 10 (47.6) | 6 (28.6) | .21 |
| Hypertension, n (%) | 13 (61.9) | 10 (47.6) | .36 |
| Cardiovascular disease, n (%) | 6 (28.6) | 4 (19.0) | .48 |
| Indication for surgery, n (%)[b] | | | |
| Symptomatic gallstone | 17 (81.0) | 15 (71.4) | .48 |
| Acute cholecystitis | 2 (9.5) | 4 (19.0) | .39 |
| Interval LC[c] after conservative treatment of acute cholecystitis | 4 (19.0) | 4 (19.0) | >.99 |
| Previous ERCP[d], n (%) | 6 (28.6) | 5 (23.8) | .73 |

[a]ASA: American Society of Anesthesiologists.

[b]Categories are not mutually exclusive.

[c]LC: laparoscopic cholecystectomy.

[d]ERCP: endoscopic retrograde cholangiopancreatography.

## Preoperative Laboratory Parameters

Preoperative laboratory findings are presented in Table 2. All laboratory values represent measurements obtained within 24 hours prior to surgery. Patients in the morphine-required group had significantly lower baseline hemoglobin (*P*=.01) and hematocrit levels (*P*=.002). Additionally, no significant between-group differences were observed in white blood cell count, neutrophil to lymphocyte ratio, liver function tests, or serum albumin levels.

**Table .** Preoperative laboratory parameters.

| Variable | Morphine required (n=21), mean (SD) | Morphine free (n=21), mean (SD) | *P* value |
|---|---|---|---|
| Hemoglobin (g/dL) | 11.95 (1.78) | 13.34 (1.60) | .01 |
| Hematocrit (%) | 35.9 (4.8) | 40.2 (4.0) | .002 |
| White blood cell count (×10³/μL) | 11.6 (19.1) | 10.7 (11.1) | .85 |
| Neutrophil to lymphocyte ratio | 6.92 (13.8) | 4.54 (6.54) | .48 |
| Aspartate aminotransferase (U/L) | 41.2 (54.4) | 45.2 (55.5) | .81 |
| Alanine aminotransferase (U/L) | 38.8 (63.7) | 46.0 (69.6) | .73 |
| Alkaline phosphatase (U/L) | 83.8 (41.2) | 95.1 (84.9) | .58 |
| Total bilirubin (mg/dL) | 0.70 (0.51) | 0.99 (1.30) | .35 |
| Serum albumin (g/dL) | 4.33 (0.39) | 4.21 (0.34) | .56 |

## Operative Details and Postoperative Outcomes

Operative and postoperative outcomes are summarized in Table 3. Mean operative time (mean 63.5, SD 15.6 minutes vs mean 58.4, SD 18.0 minutes; *P*=.32), estimated blood loss (mean 15.7, SD 11.3 mL vs mean 13.3, SD 12.5 mL; *P*=.52), and length of hospital stay (mean 2.57, SD 0.98 days vs mean 2.33, SD 0.86 days; *P*=.40) did not differ significantly between groups. Furthermore, no significant differences were observed between groups regarding trocar placement, operative technique, operative time, or intraoperative complications, and background analgesic regimens were comparable. Although patients with gallstone-related complications were included as indications for surgery, final histopathological examination of gallbladder specimens was reported as acute or chronic cholecystitis.

**Table .** Operative details, analgesic regimen, and postoperative outcomes.

| Variable | Morphine required (n=21) | Morphine free (n=21) | *P* value |
|---|---|---|---|
| Analgesic regimen | | | |
| Acetaminophen+naproxen, n (%) | 7 (33.3) | 12 (57.1) | .12 |
| Acetaminophen+tramadol, n (%) | 14 (66.7) | 9 (42.9) | .22 |
| Operative details | | | |
| Complete cholecystectomy, n (%) | 20 (95.2) | 19 (90.5) | .56 |
| Partial cholecystectomy, n (%) | 1 (4.7) | 2 (9.5) | .56 |
| Operative time (minutes), mean (SD) | 63.5 (15.6) | 58.4 (18.0) | .32 |
| Estimated blood loss (mL), mean (SD) | 15.7 (11.3) | 13.3 (12.5) | .52 |
| Postoperative outcomes | | | |
| Postoperative complication, n (%) | 1 (4.7) | 1 (4.7) | .44 |
| Nausea and vomiting, n (%) | 1 (4.7) | 1 (4.7) | .44 |
| Hospital stay (days), mean (SD) | 2.57 (0.98) | 2.33 (0.86) | .40 |
| Pathology: chronic cholecystitis, n (%) | 20 (95.3) | 17 (81.0) | .16 |
| Pathology: acute cholecystitis, n (%) | 1 (4.7) | 4 (19.0) | .30 |

## Factors Associated With Postoperative Morphine Requirement

Factors associated with postoperative morphine requirement in exploratory multivariable analysis are summarized in Table 4. Higher ASA class was associated with increased odds of morphine use within 24 hours after surgery (*P*=.01; odds ratio 6.51, 95% CI 1.37 - 30.96). Lower hemoglobin level demonstrated a trend toward association with morphine requirement but did not reach statistical significance (*P*=.07; odds ratio 0.58, 95% CI 0.32 - 1.06). Other variables, including age, sex, gallstone-related complications, and history of endoscopic retrograde cholangiopancreatography, were not significantly associated with opioid use.

**Table .** Multivariable logistic regression analysis of factors associated with postoperative morphine requirement.

| Variable | Odds ratio (95% CI) | *P* value |
|---|---|---|
| Female | 1.81 (0.26-12.80) | .55 |
| ASA[a] classification | 6.51 (1.37-30.96) | .01 |
| Hemoglobin (per g/dL) | 0.58 (0.32-1.06) | .07 |
| Age (years) | 0.98 (0.92-1.04) | .46 |
| Gallstone-related complication | 0.91 (0.16-5.15) | .91 |
| Previous ERCP[b] | 0.75 (0.12-4.56) | .75 |

[a]ASA: American Society of Anesthesiologists.

[b]ERCP: endoscopic retrograde cholangiopancreatography.

Given the limited number of outcome events, regression analyses were conducted within an exploratory framework constrained by events per variable considerations.

### Postoperative Pain Scores

Postoperative pain scores are summarized in Table 5. Patients who required morphine reported higher VAS scores during the early postoperative period, particularly at 2 hours (mean 3.29, SD 1.45 vs mean 1.93, SD 0.96; *P*=.009), 4 hours (mean 3.95, SD 1.39 vs mean 2.07, SD 1.00; *P*<.001), and 12 hours (mean 3.57, SD 1.44 vs mean 2.43, SD 1.06; *P*=.02). At 6 hours, 24 hours, and during mobilization, pain scores remained numerically higher in the morphine-required group but did not reach statistical significance.

**Table .** Comparison of postoperative pain scores between morphine-required and morphine-free groups.

| Time point | Morphine required, mean (SD) | Morphine free, mean (SD) | *P* value |
|---|---|---|---|
| 2 hours | 3.29 (1.45) | 1.93 (0.96) | .009 |
| 4 hours | 3.95 (1.39) | 2.07 (1.00) | <.001 |
| 6 hours | 3.33 (1.43) | 2.73 (1.42) | .26 |
| 12 hours | 3.57 (1.44) | 2.43 (1.06) | .02 |
| 24 hours | 2.57 (1.21) | 1.97 (0.81) | .07 |
| During mobilization | 4.29 (1.33) | 3.97 (1.25) | .40 |

These findings descriptively reflect differences in pain experience between groups and are presented to contextualize postoperative opioid requirement rather than to infer comparative analgesic effectiveness.

## Discussion

### Principal Findings

This prospective observational study suggests a clinically relevant opioid-sparing association of laparoscopic-guided subcostal TAP block in patients undergoing LC. Approximately half of the patients (n=21, 50%) did not require postoperative opioids, while those who did (n=21, 50%) received only a modest cumulative dose (mean 3.86, SD 1.39 mg).

The observed analgesic pattern was most evident during the early postoperative period, particularly between 2 and 12 hours, which is consistent with the expected pharmacodynamic profile of 0.25% bupivacaine. Within the context of this observational cohort, these findings support the feasibility of surgeon-performed TAP block as a practical adjunct to multimodal analgesia strategies, with the potential to limit postoperative opioid exposure while maintaining adequate pain control [11-14].

The results of this study are directionally consistent with previous reports suggesting that thoracoabdominal and subcostal TAP blocks are associated with improved early postoperative pain control following LC [15-17]. Importantly, this study extends the existing literature by including patients with complicated gallstone disease, a population that has been relatively underrepresented in prior research.

From a mechanistic perspective, the observed analgesic association may be attributable to localized somatic blockade of the upper abdominal wall corresponding to trocar insertion sites. Such coverage is thought to attenuate incisional and parietal peritoneal pain, which may explain the more pronounced pain relief observed during the first 12 postoperative hours [10,18]. Clinically, these observations underscore the importance of integrating regional analgesic techniques with scheduled nonopioid coanalgesics and appropriately timed rescue analgesia to maintain adequate pain control within ERAS-oriented perioperative pathways [5,6,13,14].

Exploratory predictor analysis suggested that patient-related factors were more strongly associated with postoperative opioid requirement than intraoperative variables. Higher ASA classification was independently associated with postoperative morphine use, indicating that greater comorbidity burden may be linked to increased analgesic needs despite regional blockade.

These findings support a risk-adapted approach to perioperative pain management for patients at higher risk.

Although the TAP block primarily targets somatic abdominal wall pain, unmeasured factors such as visceral pain burden, neuropathic pain components, and subtle variations in block accuracy may have influenced postoperative pain perception and opioid requirement. These factors were not objectively assessed in this study and may contribute to residual variability beyond patient-level characteristics such as ASA classification.

Beyond its analgesic association, surgeon-performed subcostal TAP block offers several practical advantages, particularly in settings with limited anesthesiology support or restricted access to ultrasound equipment. Incorporation of this technique into the laparoscopic workflow allows a consistent and equipment-independent approach to regional analgesia, aligning with broader initiatives in opioid stewardship and sustainable perioperative care [9,15-17,19].

Several limitations should be acknowledged. The single-center, nonrandomized design limits causal inference, and the absence of a control group precludes direct comparison with standard port-site local anesthetic infiltration. The modest sample size restricts statistical precision; therefore, multivariable analyses were conducted within an exploratory framework with selective variable inclusion to reduce the risk of overfitting.

In addition, no formal postoperative sensory testing was performed to objectively verify block success, primarily due to practical considerations within the perioperative workflow and the study's focus on clinically relevant outcomes. Instead, correct local anesthetic deposition was inferred from laparoscopic visualization of the Doyle bulge, indicating separation of the posterior rectus sheath and transversus abdominis muscle as a surrogate marker of appropriate fascial plane injection [20,21]. However, this visual confirmation cannot fully substitute for objective sensory testing and does not confirm the extent or consistency of dermatomal blockade.

Substitution of tramadol for nonsteroidal anti-inflammatory drugs in a small number of patients may have introduced minor confounding, although perioperative analgesic protocols were otherwise standardized. In addition, psychosocial factors known to influence postoperative pain, including anxiety, pain catastrophizing, and prior opioid exposure, were not assessed [22,23]. These limitations should be considered when interpreting the findings.

Despite these constraints, this study has several notable strengths. The prospective data collection, use of standardized perioperative analgesic pathways, and inclusion of patients with both uncomplicated and complicated gallstone disease enhance the clinical relevance of the study. Overall, the findings indicate that surgeon-performed, laparoscopic-guided subcostal TAP block is a technically straightforward and reproducible adjunct within multimodal analgesia strategies. The observed early opioid-sparing association supports the feasibility of considering this technique in ERAS-oriented perioperative pathways and broader opioid reduction efforts.

The relatively longer hospital stay observed in this cohort reflects local institutional practice, in which LC is not routinely performed as a day-case procedure. Inclusion of patients with complicated gallstone disease required preoperative admission and postoperative observation for safety. In addition, routine preoperative admission at least 1 day prior to surgery, according to institutional protocol, contributed to the overall length of hospital stay.

Future research should include multicenter randomized controlled trials comparing laparoscopic-guided TAP block with standard port-site local anesthetic infiltration. On the basis of the observed effect estimates, a sample size of approximately 80 to 100 patients per arm may provide 80% power to detect a 1-point difference in mean VAS pain score at an $\alpha$ level of .05. Incorporation of cost-effectiveness analyses, patient-reported outcome measures, and longer-term follow-up would further clarify the clinical value, scalability, and role of this technique in minimally invasive surgery.

## Conclusions

This study suggests that laparoscopic-guided subcostal TAP block may be associated with lower early postoperative pain scores and reduced opioid requirements in patients undergoing LC, including those with gallstone-related complications. Although limited by sample size, the findings support the feasibility of considering surgeon-performed subcostal TAP block as part of multimodal analgesia strategies within ERAS-oriented perioperative pathways.

## Acknowledgments

The authors thank the surgical staff at the Department of Surgery, Faculty of Medicine, Srinakharinwirot University, for their assistance with data acquisition. During the preparation of this work, the authors used ChatGPT to improve grammar and language clarity. After using this tool, the authors carefully reviewed, revised, and approved the content and take full responsibility for the published article.

## Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## Authors' Contributions

TT contributed to the conception and design of the study, data acquisition, data analysis and interpretation, and manuscript drafting. SM contributed to study design, data collection and analysis, and manuscript preparation.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Laparoscopic-guided subcostal transversus abdominis plane (TAP block).
[MP4 File, 98727 KB - periop_v9i1e87622_app1.mp4 ]

## References


1. Tullavardhana T. Critical view of safety: a safe method to prevent bile duct injury from laparoscopic cholecystectomy. J Med Health Sci 2015;22:49-57 [FREE Full text]
2. Keir A, Rhodes L, Kayal A, Khan OA. Does a transversus abdominis plane (TAP) local anaesthetic block improve pain control in patients undergoing laparoscopic cholecystectomy? A best evidence topic. Int J Surg 2013;11(9):792-794. [doi: 10.1016/j.ijsu.2013.05.039] [Medline: 23770342]
3. Law LS, Lo EA, Gan TJ, et al. Preoperative antiemetic and analgesic management. In: Gan TJ, Thacker JK, Miller TE, editors. Enhanced Recovery for Major Abdominopelvic Surgery: Professional Communications Inc; 2016:105-120.
4. Ljungqvist O, Scott M, Fearon KC. Enhanced recovery after surgery: a review. JAMA Surg 2017 Mar 1;152(3):292-298. [doi: 10.1001/jamasurg.2016.4952] [Medline: 28097305]
5. Kaye AD, Urman RD, Rappaport Y, et al. Multimodal analgesia as an essential part of enhanced recovery protocols in the ambulatory settings. J Anaesthesiol Clin Pharmacol 2019 Apr;35(Suppl 1):S40-S45. [doi: 10.4103/joacp.JOACP_51_18] [Medline: 31142958]
6. Hebbard P, Fujiwara Y, Shibata Y, Royse C. Ultrasound-guided transversus abdominis plane (TAP) block. Anaesth Intensive Care 2007 Aug;35(4):616-617. [doi: 10.1177/0310057X0703500417] [Medline: 18020088]
7. Wang W, Wang L, Gao Y. A meta-analysis of randomized controlled trials concerning the efficacy of transversus abdominis plane block for pain control after laparoscopic cholecystectomy. Front Surg 2021;8:700318. [doi: 10.3389/fsurg.2021.700318] [Medline: 34422893]
8. Elamin G, Waters PS, Hamid H, et al. Efficacy of a laparoscopically delivered transversus abdominis plane block technique during elective laparoscopic cholecystectomy: a prospective, double-blind randomized trial. J Am Coll Surg 2015 Aug;221(2):335-344. [doi: 10.1016/j.jamcollsurg.2015.03.030] [Medline: 25899736]
9. Venkatraman R, Saravanan R, Dhas M, Pushparani A. Comparison of laparoscopy-guided with ultrasound-guided subcostal transversus abdominis plane block in laparoscopic cholecystectomy - a prospective, randomised study. Indian J Anaesth 2020 Dec;64(12):1012-1017. [doi: 10.4103/ija.IJA_528_20] [Medline: 33542563]
10. Alsharari AF, Abuadas FH, Alnassrallah YS, Salihu D. Transversus abdominis plane block as a strategy for effective pain management in patients with pain during laparoscopic cholecystectomy: a systematic review. J Clin Med 2022 Nov 22;11(23):6896. [doi: 10.3390/jcm11236896] [Medline: 36498471]
11. Simpson JC, Bao X, Agarwala A. Pain Management in Enhanced Recovery after Surgery (ERAS) protocols. Clin Colon Rectal Surg 2019 Mar;32(2):121-128. [doi: 10.1055/s-0038-1676477] [Medline: 30833861]
12. Cao L, Yang T, Hou Y, Yong S, Zhou N. Efficacy and safety of different preemptive analgesia measures in pain management after laparoscopic cholecystectomy: a systematic review and network meta-analysis of randomized controlled trials. Pain Ther 2024 Dec;13(6):1471-1497. [doi: 10.1007/s40122-024-00647-w] [Medline: 39227523]
13. Barazanchi AWH, MacFater WS, Rahiri JL, et al. Evidence-based management of pain after laparoscopic cholecystectomy: a PROSPECT review update. Br J Anaesth 2018 Oct;121(4):787-803. [doi: 10.1016/j.bja.2018.06.023] [Medline: 30236241]
14. Cho HY, Hwang IE, Lee M, Kwon W, Kim WH, Lee HJ. Comparison of modified thoracoabdominal nerve block through perichondral approach and subcostal transversus abdominis plane block for pain management in laparoscopic cholecystectomy: a randomized-controlled trial. Korean J Pain 2023 Jul 1;36(3):382-391. [doi: 10.3344/kjp.23114] [Medline: 37336756]
15. Tolchard S, Davies R, Martindale S. Efficacy of the subcostal transversus abdominis plane block in laparoscopic cholecystectomy: Comparison with conventional port-site infiltration. J Anaesthesiol Clin Pharmacol 2012 Jul;28(3):339-343. [doi: 10.4103/0970-9185.98331] [Medline: 22869941]
16. Ozciftci S, Sahiner Y, Sahiner IT, Akkaya T. Is Right Unilateral Transversus Abdominis Plane (TAP) block successful in postoperative analgesia in laparoscopic cholecystectomy? Int J Clin Pract 2022;2022:2668215. [doi: 10.1155/2022/2668215] [Medline: 35685608]
17. Emile SH, Elfeki H, Elbahrawy K, Sakr A, Shalaby M. Ultrasound-guided versus laparoscopic-guided subcostal transversus abdominis plane (TAP) block versus No TAP block in laparoscopic cholecystectomy; a randomized double-blind controlled trial. Int J Surg 2022 May;101:106639. [doi: 10.1016/j.ijsu.2022.106639] [Medline: 35487422]
18. Hinrichs-Rocker A, Schulz K, Järvinen I, Lefering R, Simanski C, Neugebauer EAM. Psychosocial predictors and correlates for chronic post-surgical pain (CPSP) - a systematic review. Eur J Pain 2009 Aug;13(7):719-730. [doi: 10.1016/j.ejpain.2008.07.015] [Medline: 18952472]

19. Boonyapalanant C, Woranisarakul V, Jitpraphai S, et al. The efficacy of inside-out transversus abdominis plane block vs local infiltration before wound closure in pain management after kidney transplantation: a double-blind, randomized trial. Siriraj Med J 2022;74(4):233-238. [doi: 10.33192/Smj.2022.29]
20. Vindal A, Sarda H, Lal P. Laparoscopically guided transversus abdominis plane block offers better pain relief after laparoscopic cholecystectomy: results of a triple blind randomized controlled trial. Surg Endosc 2021 Apr;35(4):1713-1721. [doi: 10.1007/s00464-020-07558-9] [Medline: 32297054]
21. Suragul W, Tantawanit A, Rungsakulkij N, et al. Effect of local anaesthetic infiltration on postoperative pain after laparoscopic cholecystectomy: randomized clinical trial. BJS Open 2022 May 2;6(3):zrac066. [doi: 10.1093/bjsopen/zrac066] [Medline: 35639946]
22. Ip HYV, Abrishami A, Peng PWH, Wong J, Chung F. Predictors of postoperative pain and analgesic consumption: a qualitative systematic review. Anesthesiology 2009 Sep;111(3):657-677. [doi: 10.1097/ALN.0b013e3181aae87a] [Medline: 19672167]
23. Riddle DL, Wade JB, Jiranek WA, Kong X. Preoperative pain catastrophizing predicts pain outcome after knee arthroplasty. Clin Orthop Relat Res 2010 Mar;468(3):798-806. [doi: 10.1007/s11999-009-0963-y] [Medline: 19585177]

## Abbreviations

**ERAS:** enhanced recovery after surgery
**LC:** laparoscopic cholecystectomy
**TAP:** transversus abdominis plane
**VAS:** Visual Analog Scale

*Edited by N Rohatgi, S Pan; submitted 11.Nov.2025; peer-reviewed by A Jahangeer, ADK Hill, H Abdallah; revised version received 22.Feb.2026; accepted 25.Feb.2026; published 25.Mar.2026.*

*Please cite as:*
*Mahasupachai S, Tullavardhana T*
*Clinical Feasibility and Outcomes of Surgeon-Performed Laparoscopic-Guided Subcostal Transversus Abdominis Plane Block in Laparoscopic Cholecystectomy: Prospective Observational Study*
*JMIR Perioper Med 2026;9:e87622*
*URL: https://periop.jmir.org/2026/1/e87622*
*doi:10.2196/87622*

# Immersive Virtual Reality for Pain and Relaxation in Older Adults Following Elective Inpatient Abdominal Surgery: Single-Arm Study Examining Feasibility and Acceptability

Christina Keny[1], PhD, MS, RN; Ujala Shafiq[1], MS; Karl Lorenz[2], MD, MSHS; Marcia Russell[3], MD; Heather Leutwyler[1], PhD, NP; Laura M Wagner[1], PhD, NP; Victoria Tang[4], MD, MAS; Linda G Park[1], PhD, NP

[1]Department of Community Health Systems, School of Nursing, University of California, San Francisco, Box 0608, 490 Illinois Street, #93P, San Francisco, CA, United States

[2]Department of Primary Care and Population Health (PCPH), Stanford University School of Medicine, Palo Alto, CA, United States

[3]Department of Surgery, David Geffen School of Medicine at UCLA, Los Angeles, CA, United States

[4]Department of Medicine, Division of Geriatrics, University of California, San Francisco, San Francisco, CA, United States

**Corresponding Author:**
Christina Keny, PhD, MS, RN
Department of Community Health Systems, School of Nursing, University of California, San Francisco, Box 0608, 490 Illinois Street, #93P, San Francisco, CA, United States

## Abstract

**Background:** There is mounting evidence to suggest that immersive virtual reality (IVR) can improve pain in older adults in community settings, yet the use of IVR postoperatively in the acute postoperative period following major elective abdominal surgery remains largely underexplored.

**Objective:** This single-arm pilot study aimed to assess the feasibility, acceptability, and preliminary impact of IVR on self-reported postoperative pain and relaxation levels in older adults following elective major abdominal surgery.

**Methods:** We recruited individuals aged 55 years and older undergoing elective abdominal surgery at an academic medical center from October 2023 to February 2024. We evaluated feasibility through accrual rate, intervention completion, and questionnaire compliance; acceptability via the System Usability Scale (SUS) and a user experience survey; and tolerability by monitoring self-reported side effects. The preliminary impact of IVR on self-reported pain intensity and relaxation levels was assessed through pre- and postintervention comparisons.

**Results:** A total of 29 participants, with a median age of 73 (IQR 55 - 81) years, were enrolled and completed at least 1 IVR session, with 19 also completing a second session. Perceived usability and overall acceptance of IVR were high, with minimal side effects reported. In terms of the preliminary impact of IVR, statistically significant improvements were observed in both pain and relaxation levels from pre- to post-IVR on day 1 and day 2.

**Conclusions:** This study suggests the feasibility and acceptability of IVR as a potential future intervention for postoperative pain management and enhancing relaxation among older adults following elective inpatient abdominal surgery. The preliminary findings suggest the need for large-scale studies across additional complex inpatient abdominal surgeries to confirm the acceptance and efficacy of IVR as a postoperative pain management intervention across a wide range of diverse older demographics. Future research is critical to evaluating the therapeutic potential of IVR in a variety of surgical and patient-specific contexts.

**Trial Registration:** ClinicalTrials.gov NCT06095661; https://clinicaltrials.gov/study/Nct06095661

***(JMIR Perioper Med 2026;9:e81791)*** doi:10.2196/81791

**KEYWORDS**

geriatrics; surgery; pain; virtual reality; postoperative pain; geriatric surgery

## Introduction

Nearly 4 million operations are performed annually on individuals aged 65 years or older in the United States, a number that is expected to rise significantly as the population continues to age [1,2]. Optimal pain management following major surgery is crucial for older adults who face unique risks associated with uncontrolled postoperative pain, such as delirium, functional decline, and reduced psychosocial well-being [3-6]. Effectively managing pain in older surgical adults is often complicated by challenges stemming from age-related physiological changes, the presence of multiple comorbidities, and the intricacies of polypharmacy [7,8]. Moreover, older adults have an increased risk of developing opioid-related adverse events, addiction,

and/or chronic pain following surgery when compared to younger age groups [9-12]. These factors may contribute to the heightened risks associated with traditional pharmacologic pain management in the older adult [13-15].

Given the complexities and inherent risks in managing acute postoperative pain in this demographic, there is a growing interest in exploring innovative nonpharmacological methods for pain [16-18]. Among these emerging solutions, immersive virtual reality (IVR) has garnered significant interest from both the clinical and research communities [19]. Defined by the use of a head-mounted display (HMD) with motion tracking capabilities, IVR effectively provides users with a believable sense of reality while engaged in a virtual environment [20]. This profound sense of "presence" within the virtual environment often provides positive distraction away from pain stimuli [21-23].

Research on IVR has shown significant potential in reducing acute pain across various clinical settings, such as burn wound care [24,25]. Recent studies have also demonstrated that IVR is effective in reducing postoperative pain in pediatric and young to middle-aged adults following various surgical procedures [25,26]. Contrary to the common belief that older adults are hesitant to embrace new technologies, there is mounting evidence to suggest that using IVR for pain management in the older adult demographic is promising [27-30]. Research focused on IVR use for chronic pain management in community-dwelling older adults has demonstrated improved pain tolerance with a high degree of acceptance as part of pain management [31,32]. Finally, studies have also indicated the acceptability and efficacy of IVR use for postoperative pain among older adults, specifically in elective total knee arthroplasty operations [33,34].

It is key to recognize that the user experience of IVR for postoperative pain management may differ across various older adult subgroups and types of surgical procedures [30,35]. Appreciating these distinctions is vital for tailoring IVR applications to effectively manage pain in a wide range of surgical scenarios and across different older age demographics. Thus, IVR could serve as a promising alternative or adjunct to traditional pain management techniques. However, the feasibility, acceptability, and tolerability of IVR for postoperative pain among older adults across a spectrum of major surgical procedures, including complex abdominal operations, remains largely underexplored. Therefore, the aim of this single-arm study was to investigate the initial feasibility and acceptability (primary outcome), and the preliminary impact (secondary outcome) of IVR on postoperative pain and relaxation levels in older adults during the initial days following inpatient elective abdominal surgery.

## Methods

### Study Design

This study used a prospective pretest-posttest single-arm study design to ascertain the feasibility, acceptability, and preliminary impact of IVR on pain and relaxation outcomes among older adults following elective major abdominal surgery. The participant enrollment sample size was set for practical reasons and not driven by power analysis for this initial study [36]. This study followed the CONSORT (Consolidated Standards of Reporting Trials) extension to pilot and feasibility studies statement, which is recommended for adaptation in nonrandomized feasibility studies [37]. This study design also aligns with the recommendations and methodological framework proposed by the Virtual Reality Clinical Outcomes Research Experts (VR-CORE) on best practices for the development and testing of IVR treatments in clinical care [38].

### Ethical Considerations

The study received approval from the University of California, San Francisco Institutional Review Board (IRB #19 - 28391) and is registered under clinicaltrials.gov (NCT06095661). All participants were provided with written and verbal informed consent before taking part in the study. Data were stored on secure, password-protected servers, and all analyses used deidentified data in accordance with the approved protocol. All responses to study questionnaires were deidentified and entered directly into an electronic tablet that was password-protected. A one-time US $25 gift card was provided to all participants for their participation in at least 1 IVR session. A subgroup of participants was additionally offered the option to complete a single user experience survey prior to hospital discharge. Those who opted to participate in the survey received an additional US $25 gift card.

### Participants

We used purposeful sampling to recruit adults aged ≥55 years undergoing elective inpatient abdominal surgery at the University of California, San Francisco, Colorectal and General Surgery clinics. Those who were potentially eligible for the study and who were interested in participation were screened via telephone. Inclusion criteria were individuals anticipated to have an elective abdominal operation requiring hospitalization for at least 48 hours after surgery and those who were able to speak and write in English. Exclusion criteria encompassed individuals with a reported history of self-reported motion sickness, severe cognitive impairment, epilepsy, eye, neck, or face injuries, blindness or severe visual impairment, severe hearing loss, or acute illness hindering postsurgery IVR use. Participants with immediate preintervention nausea, vomiting, or dizziness were also excluded. All participants continued to receive their usual surgical care as per the recommendations of clinical providers and were not asked to decline or change any adjunct strategies for pain management, as the intent of the study was to understand the initial feasibility and acceptability of IVR during the first days after surgery on an inpatient hospital unit.

### Intervention

This study used the REAL System i-Series IVR HMD from Penumbra, Inc, featuring built-in audio and gaze-controlled navigation [39]. This IVR system, preloaded with various 360-degree immersive environments for positive distraction and relaxation, includes experiences such as mindful meditation, travel, nature scenes, and games. The IVR system provides motion tracking through sensors embedded in the headset,

capturing all possible participant movements, thus facilitating an extensive immersive experience. The decision to leverage this specific IVR device was 2-fold: (1) the HMD unit was preloaded with a built-in library of experiences, allowing the user access to a wide variety of IVR-positive distraction environments, and (2) the device offered gaze-controlled navigation, potentially allowing ease of use in the immediate postoperative phase of care following major abdominal surgery.

## Procedure

During the preoperative surgery visit, clinical staff provided possible participants with an informational flyer and an email introducing the study. Participant information was gathered from the electronic medical record, and potential participants were then contacted by phone and screened for eligibility by the research team. Screening included assessment of cognitive function using the Short Portable Mental Status Questionnaire (SPMSQ), self-reported history of motion sickness, epilepsy, blindness, severe hearing deficits, and any current eye, face, or neck injuries. If deemed eligible and interested in study participation following the initial screening, participants electronically received the informed consent form. Once the consent form was signed, participants were asked to electronically complete an online questionnaire for sociodemographic and clinical data prior to their date of surgery. Prior to surgery, participants received an instructional video link demonstrating the use of the IVR headset, along with a catalog of available immersive content options. Immediately before each intervention session, the study team again reviewed headset use and content selection with the participant. A trained study team member with expertise in the IVR protocol remained present throughout the session to assist with setup, navigation, and troubleshooting as needed.

All participants enrolled in the study were provided with the opportunity to engage in at least 1 IVR session within their hospital room. These sessions were made available starting from the next day following surgery and could extend up to the second day after surgery, ensuring a maximum offering of 2 IVR sessions in total. The IVR intervention was administered to the patient in a seated or lying position by a member of the research team, who was present during and up to 15 minutes after each IVR session. Participants then choose their desired experience within the IVR content library. IVR program preference selection and length of the session were determined by the participant, up to a maximum of 30 minutes per session.

Immediately before and after the IVR intervention, participants reported their pain intensity level and state of relaxation on an 11-point Numeric Rating Scale (NRS) ranging from “0” representing “no pain” or “not relaxed at all” to “10” representing “pain as bad as you can imagine” or “as relaxed as possible.” Adverse outcomes were assessed up through the first 15 minutes after each IVR session using an adapted 4-item Simulator Sickness Questionnaire (SSQ) [40]. All responses were deidentified and entered directly into an electronic tablet that was password-protected.

## Outcome Measures

### *Acceptability*

In the context of this study, acceptability refers to participants’ willingness to use IVR in the initial 2 days of their hospitalization and their ability to tolerate IVR use, with minimal side effects reported. Acceptability was assessed by the System Usability Scale (SUS) in all participants (N=29), with a subgroup of 21 participants also completing a user experience survey. After at least 1 session of IVR, all participants were asked to complete the SUS once. SUS is considered a valid and reliable instrument for measuring perceived usability of a technology system and consists of a 10-item questionnaire with 5 response options, with “1” representing “strongly disagree” and “5” representing “strongly agree” [41]. A higher score indicates greater self-reported usability, reflecting a positive attitude toward using the system [42]. An 8-item user experience survey created by the research team using a 5-point Likert scale, ranging from 1 (“totally disagree”) to 5 (“totally agree”), was also administered to quantify participant satisfaction with IVR. The scores from the user experience survey for each sentiment level (strongly disagree-1, disagree-2, neutral-3, agree-4, and strongly agree-5) are presented as a mean and SD. Tolerability refers to the evaluation of adverse events that occurred as a result of IVR use, related to either the hardware or software components [38]. Adverse outcomes, which include symptoms such as nausea, headache, blurred vision, and dizziness, were assessed using a 4-item adapted questionnaire of the SSQ [40]. This questionnaire was administered immediately after each IVR session, allowing participants to indicate the presence or absence of these symptoms with a “Yes” or “No” response. Participants also had the option to free-text any side effects they felt occurred during or immediately after IVR use.

### *Feasibility*

In this study, feasibility is defined as the extent to which potentially eligible participants consented to join the study during the recruitment phase and the degree to which enrolled participants successfully completed the IVR intervention and all questionnaires. To evaluate feasibility, we measured the rate of participant accrual, reasons for nonparticipation, the successful completion of the intervention on the first and second days after surgery, and the mean duration time spent using IVR during each session. We also captured reasons for not completing a second IVR session.

We also evaluated the feasibility by the rate at which participants completed baseline questionnaires. The baseline characteristics captured through questionnaires were self-reported perceived health status, anxiety, depression, and pain catastrophizing prior to the operation. Perceived health status was measured using the EQ-5D-5L questionnaires’ Visual Analog Scale (VAS), with scores ranging from 0 to 100, with a higher score indicating higher perceived health [43,44]. Anxiety was measured using the Generalized Anxiety Disorder-7 (GAD-7) scale, with higher scores indicating higher anxiety levels (total score for GAD is 0 - 21) [45]. Depression was assessed using the Patient Health Questionnaire-8 (PHQ-8), which ranges from 0 to 24, with a higher score indicating higher levels of depression [46]. Finally,

we leveraged the Pain Catastrophizing Scale (PCS), where higher scores signify more intense negative thoughts and feelings toward pain [47].

### *Preliminary Clinical Impact on Pain and Relaxation*

Preliminary clinical impact of IVR on pain intensity levels and state of relaxation was measured through pre- and postintervention mean differences using independent paired sample *t* tests if the data were deemed normal. Nonparametric continuous data were evaluated using the Wilcoxon signed-rank test. Pain intensity level was measured on an 11-point NRS and ranged from "0" representing "no pain" to "10" representing extreme pain. State of relaxation is also measured on an 11-point NRS with "0" representing "not relaxed at all" to "10" representing "as relaxed as one could imagine." Both pain and relaxation were assessed immediately prior to and after each IVR session.

### Data Analysis

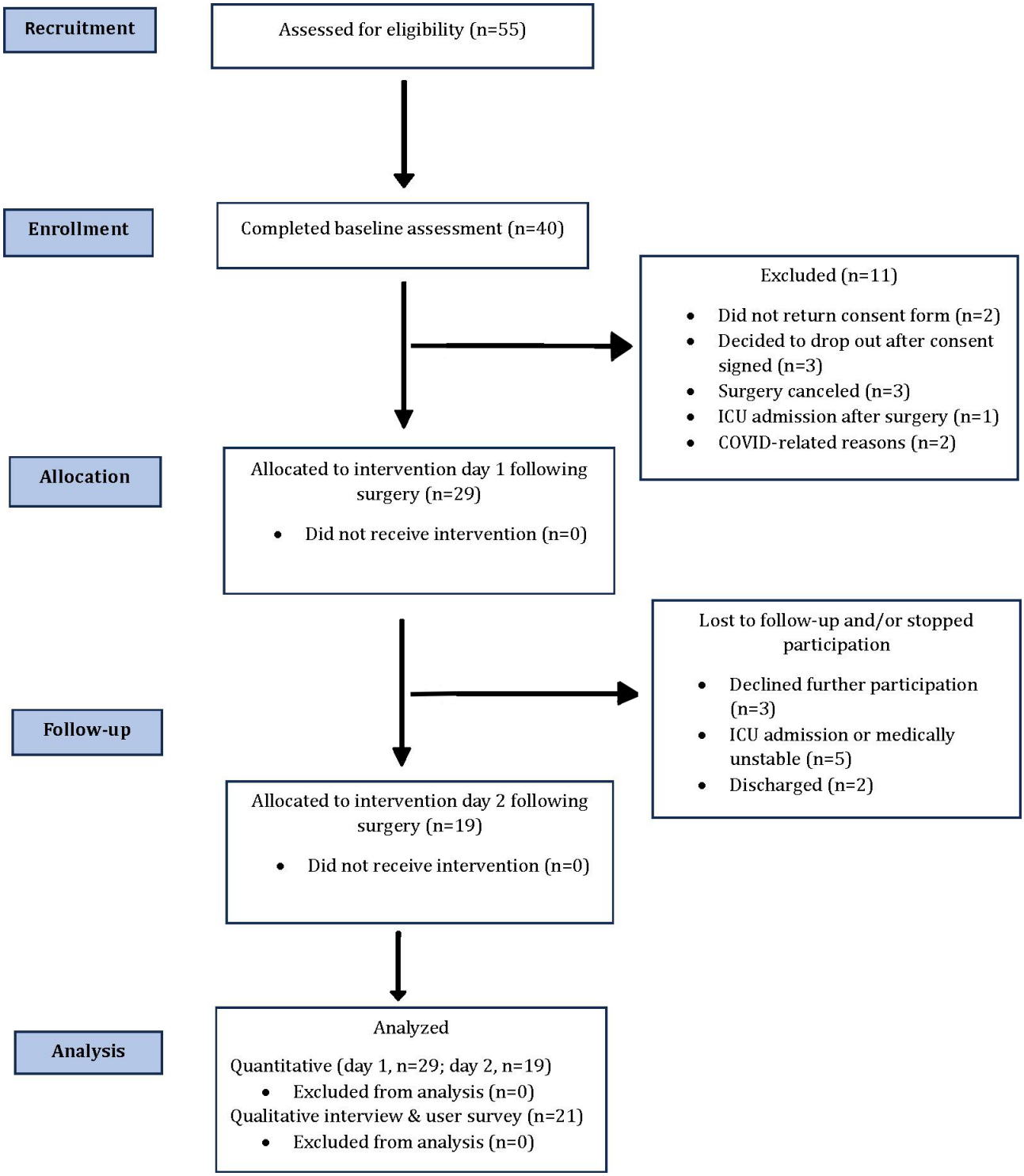
Acceptability, feasibility, and tolerability are reported as descriptive statistics. User experience surveys for each sentiment level were presented as a mean and SD. Prior to analyzing the preliminary effects of the IVR intervention, we assessed the normality of the distribution of pain and relaxation levels using the Shapiro-Wilk test. Normal data were compared as pre-post mean differences with independent paired 2-sample *t* tests. Nonparametric data were reported as a median and compared using the Wilcoxon signed-rank test. The significance level was set at $P<.05$. All statistical analyses were performed using the STATA statistical software, version 18 SE (StataCorp LLC) [48].

## Results

### Characteristics of Participants

Fifty-five participants scheduled for elective inpatient abdominal surgery were assessed for study eligibility between October 2023 and February 2024 (Figure 1). Among possible participants, 15 in total were excluded due to not meeting the inclusion criteria. A total of 40 participants completed the baseline questionnaires prior to surgery. Of the 40 participants enrolled, 11 withdrew for several reasons, including failure to complete the written informed consent prior to surgery (n=2). Postoperative consent was not pursued in these cases because participants had undergone general anesthesia. Additional reasons for withdrawal included surgery cancellation (n=3), postoperative Intensive Care Unit admission (n=1), and testing positive for COVID-19 following surgery (n=2). Three additional participants withdrew after providing consent for unknown reasons. A total of 29 enrolled participants were allocated to and completed the first IVR intervention the next day following their surgery, with 19 (65.5%) additionally completing a second intervention the next day. Most participants reported no prior experience with IVR (27/29, 93.1%).

**Figure 1.** CONSORT (Consolidated Standards of Reporting Trials) flow diagram.




Among 29 participants, the median age was 73 (IQR 55-81) years. A total of 79.3% (23/29) identified as White, 62.10% (18/29) were female, and all but 2 participants reported at least some level of college education (27/29, 93.1%; Table 1). Nearly half of the participants (14/29, 48.3%) underwent a low anterior resection abdominal operation, with cancer as the most common indication for surgery (19/29, 65.5%). More than half of participants reported pain in the 2 weeks prior to surgery (16/29, 55.2%), with a few individuals taking opioid medications for pain (n=3). In our evaluation of the baseline characteristics of the study sample, we found that participants typically described their perceived health status as generally good. Additionally, they reported experiencing mild levels of anxiety and depression, alongside minimal tendencies toward pain catastrophizing, before undergoing surgery and the IVR intervention (Table 1).

**Table .** Sociodemographic, baseline characteristics, and clinical descriptives (N=29).

| Characteristics | Values |
|---|---|
| Age (years), median (range) | 73 (55 - 81) |
| Sex, n (%) | |
| Male | 11 (37.9) |
| Female | 18 (62.1) |
| Race or ethnicity, n (%) | |
| White or Caucasian | 23 (79.3) |
| Asian | 4 (13.8) |
| Hispanic | 2 (6.9) |
| Black or African American | 0 (0) |
| Relationship status, n (%) | |
| Currently married | 14 (48.3) |
| Divorced | 1 (3.4) |
| Single | 13 (44.8) |
| Widowed | 1 (3.4) |
| Level of education, n (%) | |
| High school diploma | 2 (6.9) |
| Some college, no degree | 5 (17.2) |
| Any college, graduate, or professional degree | 22 (75.9) |
| Primary indication for surgery, n (%) | |
| Cancer | 19 (65.5) |
| Primary types of abdominal procedures, n (%) | |
| Laparoscopic low anterior resection | 11 (37.9) |
| Robotic-assisted low anterior resection | 3 (10.3) |
| Laparoscopic colectomy | 5 (17.2) |
| Open colectomy | 2 (6.8) |
| Open colostomy revision or takedown | 3 (10.3) |
| Ileostomy takedown | 3 (10.3) |
| Open abdominal perineal resection | 1 (3.4) |
| Robotic-assisted rectopexy | 1 (3.4) |
| Prior pain and virtual reality use, n (%) | |
| Pain in the past week prior to surgery | 16 (55.2) |
| Current opioid use for pain prior to surgery | 3 (10.3) |
| Prior virtual reality use | 2 (6.9) |
| Baseline scores, mean (SD) | |
| Pain Catastrophizing Scale (PCS)[a] | 14.56 (14.21) |
| Health State Questionnaire – EQ-5D-5L[b] | 71.10 (22.3) |
| Generalized Anxiety Disorder 7-item (GAD-7)[c] | 5.56 (5.9) |
| Patient Health Questionnaire-8 (PHQ-8)[d] | 5.03 (5.5) |

[a]Pain Catastrophizing Scale consists of 13 items, with each item on a scale from 0 to 4 based on their thoughts when experiencing pain. The total score can range from 0 to 52, with higher scores indicating greater levels of pain catastrophizing.

[b]The EuroQol 5-Dimension 5-Level questionnaire uses a Visual Analog Scale (VAS), where the end points are labeled as the “best imaginable health state” and the “worst imaginable health state.” The VAS score ranges from 0 to 100, with the higher score indicating higher perceived health.

[c]Generalized anxiety disorder 7-item scale is a self-reported questionnaire used to assess the severity of anxiety symptoms with each item scored from 0 (not at all) to 3 (nearly every day). The total score ranges from 0 to 21, with higher scores indicating higher anxiety levels.

[d]The Patient Health Questionnaire-8 measures the severity of depressive symptoms. Each item is scored on a scale from 0 to 3. The sum of all items is a range of 0 to 24, with a higher score indicating greater levels of depressive symptoms.

## Acceptability

The mean duration of IVR use in the initial session for 29 participants was 19.14 (SD 7.67) minutes. Moreover, 19 participants additionally completed a second session, during which the mean usage time was 16.78 (SD 6.13) minutes (Table 2). The most common IVR experiences chosen by study participants were guided travel, followed by mindfulness and meditation, with nearly half of the participants choosing more than 1 IVR experience during a single session. The results indicated high perceived usability of IVR in this sample as demonstrated by a high mean SUS score of 88.10 (SD 6.15). SUS scores above 68 are considered above average and are an indicator of good usability [41]. Individual adjusted raw mean scores for each SUS item were generally >3, indicating positive usability for each SUS item statement (adjusted items ranged from 0 to 4, with 4 as more desirable per each item). Higher scores after adjustment indicate better usability (Table 3).

**Table .** Mean time spent using immersive virtual reality and content selection.

| Usage time and content selection | 1 day after surgery (N=29) | 2 days after surgery (n=19) |
|---|---|---|
| Time spent in IVR[a] (min), mean (SD; range) | 19.14 (7.67; 6-30) | 16.78 (6.13; 3-30) |
| Participant IVR content selection[b], n (%) | | |
| Guided travel | 20 (68.9) | 10 (52.6) |
| Mindfulness and meditation | 7 (24.1) | 5 (26.3) |
| Arctic cold and/or underwater | 4 (13.7) | 3 (15.8) |
| Forests and/or wildlife | 2 (6.8) | 1 (5.3) |
| Games | 1 (3.4) | 1 (5.3) |

[a]IVR: immersive virtual reality.

[b]Participants had the option to choose as many experiences as desired, up to a maximum of 30 minutes of use per session, in any of the content categories offered within the preloaded software library.

**Table .** System Usability Scale (SUS) ratings.

| SUS[a] adjusted raw score per item (N=29) | Score, mean (SD) | Range |
|---|---|---|
| I think that I would like to use this system frequently. | 3.5 (0.58) | 2 - 4 |
| I found the system unnecessarily complex. | 3.8 (0.77) | 0 - 4 |
| I thought the system was easy to use. | 3.7 (0.47) | 3 - 4 |
| I think that I would need the support of a technical person to be able to use this system. | 3.2 (0.85) | 1 - 4 |
| I found the various functions in this system were well integrated. | 2.48 (0.87) | 1 - 4 |
| I thought there was too much inconsistency in this system. | 3.7 (0.66) | 2 - 4 |
| I would imagine that most people would learn to use this system very quickly. | 3.6 (0.50) | 3 - 4 |
| I found the system very cumbersome to use. | 3.9 (0.37) | 2 - 4 |
| I felt very confident using the system. | 3.7 (0.47) | 3 - 4 |
| I needed to learn a lot of things before I could get going with this system. | 3.8 (0.41) | 3 - 4 |

[a]SUS: System Usability Scale.

The original responses are given on a Likert scale from 1 (strongly disagree) to 5 (strongly agree) for each of the 10 items. After adjusting the scale of negatively worded questions (items 2, 4, 6, 8, and 10) by subtracting their scores from 5 and for positively worded questions (items 1, 3, 5, 7, and 9) obtained by subtracting 1, adjusted scores will range from 0 to 4. After adjustment, “0” represents a negative usability experience as related to the item statement and “4” represents a positive usability experience for each item. Higher scores after

adjustment indicate better usability (adapted from the study by Brooke [42]).

Study participants reported an overall positive experience with using IVR as indicated in the survey questionnaire (Table 4). All study participants marked responses of "agree-4" or "strongly agree-5" to the statement "I liked the virtual reality experience." Most of the study participants (27/29, 93.1%) also marked "agree-4" or "strongly agree-5" that IVR improved their postoperative pain. Furthermore, most participants agreed that they would use IVR again for pain (26/29, 89.6%) or anxiety (25/29, 86.2%). Finally, all participants (N=29) marked "strongly agree" to the statement, "I would recommend virtual reality to other older surgical patients."

**Table .** User experience survey responses (N=29).[a]

| User experience item | Scores, mean (SD) |
|---|---|
| Would recommend VR[b] to others | 5.0 (0.0) |
| Would use VR again for pain | 4.8 (0.4) |
| I liked the VR experience | 4.8 (0.7) |
| Would use VR again for anxiety | 4.7 (0.7) |
| The audio sound was pleasant | 4.6 (0.6) |
| VR improved my pain | 4.5 (0.8) |
| The image quality was pleasant | 3.8 (0.8) |
| The headset was comfortable | 3.1 (1.4) |

[a]Participants rated each statement on a 5-point Likert scale (1=totally disagree, 2=disagree, 3=neutral, 4=agree, 5=totally agree). Mean scores and SDs are reported for each item and are ordered from highest to lowest mean rating.

[b]VR: virtual reality.

Nearly all participants (28/29, 96.6%) completed one or more IVR sessions without self-reported side effects (eg, dizziness, headache, eye strain, and nausea). One participant reported mild face and chest skin redness that occurred nearly 24 hours after the first IVR session. Upon further investigation, it was unclear as to the exact cause of the skin irritation, whether due to the IVR headset foam padding or related to recent medications as part of surgical care. Although unlikely related to the use of IVR equipment, the possible adverse event was reported out of an abundance of caution.

## Feasibility

To evaluate overall feasibility, we measured the rate of accrual and reasons for nonparticipation, as well as successful completion of IVR on the first and second day after inpatient abdominal surgery (Figure 1). Of the 55 potential participants assessed for eligibility, 7 individuals did not meet inclusion requirements during telephone screening, 7 declined to participate after learning more about the study, stating extreme anxiety over surgery (n=5) or fear of new technology (n=2) as the main reasons for nonparticipation, and 1 individual refused to state a reason for declining participation (Figure 1). The remaining participants (n=40) were deemed eligible and agreed to participate. Of the 40 participants, a total of 11 either dropped out or were excluded before the intervention was administered due to the following reasons: did not return the consent forms (n=2), decided to drop out of the study prior to the date of surgery for unspecified reasons (n=3), surgery was canceled (n=3), Intensive Care Unit admission directly from the operating room (n=1), and COVID-19–related reasons (n=2).

A total of 29 enrolled participants allocated to the intervention on the day following surgery all completed the first IVR session, with 19 additionally completing the second IVR session the next day. The most common reason for not completing a second IVR session the next day was typically due to a change in complexity of care or severe nausea or vomiting not related to IVR (n=5). All participants allocated to the first and second IVR sessions completed all baseline questionnaires, as well as all pre- and postintervention questionnaires, with no missing or unanswered items.

## Preliminary Clinical Impact of IVR on Pain and Relaxation

Significant improvements were observed in both pain and relaxation levels from pre- to post-IVR on both day 1 and day 2 following surgery (Table 5). The preliminary impact of IVR on pain levels was analyzed using a paired 2-sample *t* test. On day 1 following surgery, post-IVR mean pain levels showed a significant reduction as compared to pre-IVR pain levels, with a mean improvement of 2.65 (SD 2.0), representing approximately a 50% reduction in pain scores (95% CI 1.89-3.41; $P<.001$). Similarly, on day 2 following surgery, results indicated a significant decrease in pain levels from pre- to post-IVR, with a mean pain level decrease of 2.05 (SD 1.5; 95% CI 1.33-2.78; $P<.001$).

**Table .** Pre-to-post immersive virtual reality changes in self-reported pain and relaxation.

| Pre-post IVR[a] pain and relaxation levels | Mean (SD) or median (IQR) | 95% CI | Test statistic | *P* value |
|---|---|---|---|---|
| Pain level day 1, N=29, mean (SD) | | | N/A[b] | .001 |
| Pre-IVR | 5.17 (2.1) | 4.37 - 5.97 | | |
| Post-IVR | 2.51 (1.5) | 1.91 - 3.16 | | |
| Change | 2.65 (2.0) | 1.89 - 3.41 | | |
| Pain level day 2, N=19, mean (SD) | | | N/A | .001 |
| Pre-IVR | 4.84 (1.6) | 4.08 - 5.60 | | |
| Post-IVR | 2.79 (1.5) | 2.06 - 3.51 | | |
| Change | 2.05 (1.5) | 1.33 - 2.78 | | |
| Relaxation level day 1, N=29, median (IQR) | | | Z=−4.6 | <.01 |
| Pre-IVR | 3 (3-6) | | | |
| Post-IVR | 8 (7-9) | | | |
| Change | __[c] | | | |
| Relaxation level day 2, n=19, median (IQR) | | | Z=−3.85 | <.01 |
| Pre-IVR | 4 (4-5) | | | |
| Post-IVR | 8 (7-9) | | | |
| Change | — | | | |

[a]IVR: immersive virtual reality.

[b]N/A: not applicable.

[c]Not available.

Finally, participants reported a significantly higher level of relaxation immediately following IVR as compared to pre-IVR relaxation levels on both day 1 and day 2 following surgery. The Wilcoxon signed-rank test, applied due to the nonnormal distribution of relaxation scores, revealed significant findings. One day following surgery, the median relaxation score prior to using IVR was 3 (IQR 3-6), which increased to a median score of 8 (IQR 7-9) immediately following IVR usage, indicating a statistically significant improvement in relaxation levels from pre- to postintervention (Z=−4.6; *P*<.01). On the second day after surgery, the median score before intervention was 4 (IQR 4-5), which increased to a median relaxation score of 8 (IQR 7-9) following the intervention, demonstrating a significant improvement in relaxation scores from pre- to post-IVR use (Z=–3.85; *P*<.01).

# Discussion

## Principal Findings

Our study indicates that IVR is a feasible, acceptable, and well-tolerated intervention for postoperative pain management and relaxation in older adults in the initial days following elective major inpatient abdominal surgery. Participants reported high perceived usability and acceptance, with minimal side effects. IVR use resulted in statistically significant reductions in self-reported pain and improvements in relaxation from pre- to postintervention on both day 1 and day 2, with most participants indicating a willingness to use IVR again and recommend it to other older surgical patients. While prior studies have established the feasibility of IVR for preoperative anxiety and postoperative exercise in older adults undergoing colorectal surgery [49,50], our study is among the few to extend these findings to postoperative pain and relaxation across a wider range of elective abdominal procedures. These results support the potential of IVR as a nonpharmacologic option for pain management and relaxation in this population.

## Acceptability, Feasibility, and Tolerability

Our research both corroborates and diverges from existing literature on the application of IVR for managing pain and facilitating relaxation. Recent meta-analyses have suggested that IVR is efficacious in reducing acute pain following a variety of surgical and medical procedures [25,51]. However, except for total knee arthroplasty operations, there is a notable gap in research regarding the acceptability, feasibility, and tolerability of IVR among older surgical adults in the initial days following various types of major elective inpatient surgeries [52]. Some studies suggest that hospitalized older adults older than the age of 60 are more likely to decline participation in IVR studies, related to a lack of understanding or perceived usefulness of IVR [53,54]. Other research indicates a possibly lower acceptance of IVR among older adults in acute care settings due to negative attitudes and anxiety toward using technology [52,55]. Furthermore, except for those who developed postoperative clinical complications precluding the use of IVR, the dropout rate among our study participants was notably low at 12.5% (n=5). In both community and residential contexts, as well as within the scope of colorectal cancer care, existing

research has indicated that the use of IVR as a positive distraction is both feasible and highly accepted among older adults for alleviating chronic pain in nonhospital settings [31,56-58]. Moreover, studies on the adoption of emerging technologies such as IVR suggest that older adults are generally more receptive to using new technology when they are offered a broad range of choices and the autonomy to select content and information according to their unique preferences [59-63].

Regarding the tolerability of IVR use in older adults, there have been minor reports of motion sickness and occasional discomfort with the IVR headset noted in prior research [52,64,65]. In our study, there were no reports of motion sickness during or after use. This finding may have been due to limiting each IVR session to a maximum of 30 minutes. A recent review of IVR use in older adults across a wide spectrum of nonsurgical settings found cybersickness (eg, dizziness, nausea, eye strain, etc) to be minimal across 39 studies [66,67]. Also related to the tolerability of IVR, in our study, 1 participant reported skin irritation, an adverse effect for which there is little to no existing research quantifying its incidence rate [68]. Finally, our user experience survey results suggest that while the overall satisfaction with the IVR system was positive, a significant number of participants expressed discomfort with the headset. This reported discomfort is consistent with prior research indicating that while older adults in community settings have found the headsets to be bulky and uncomfortable, their overall experience with IVR was enjoyable [30].

## Clinical Implications

The integration of IVR into the care of older adults following major abdominal surgery presents a promising avenue for enhancing pain management and relaxation strategies, with the potential to improve patient-reported outcomes, patient satisfaction, and the overall recovery experience. First, despite existing concerns about older adults' willingness to engage with new technologies, our study found a high level of acceptability and tolerability of IVR for pain and relaxation. This finding suggests that with proper introduction and support, as we provided in our study, older adults might be open to using innovative technology such as IVR in the acute care setting as part of their care [28,69]. Second, our results hold particular significance in the surgical care of older adults, where IVR could act as a cost-effective and safe alternative or adjunct to conventional pain management approaches [9,70,71]. Some studies indicate that IVR retains the possibility to lessen opioid medication usage and its associated side effects [72,73]. In the last decade, government agencies and clinical professional organizations, including the American College of Surgeons' Geriatric Surgery Verification program, have increased emphasis on the importance of incorporating opioid-sparing methods into pain treatment strategies, including nonpharmacological interventions, especially in vulnerable groups such as older adults [16,74]. The prominence of this push can also be exemplified through a recent ruling by the Centers for Medicare & Medicaid Services, indicating that one type of an integrated software or hardware IVR device may be eligible for Medicare insurance coverage [75]. In the near future, it is expected that improved insurance coverage will greatly increase the accessibility of such technologies for older adults as part of their pain management.

Furthermore, introducing IVR into the care of older adults may initially be met with uncertainty, by both clinicians and older adults, due to prevailing beliefs about technological proficiency or appropriateness in this age group. This view may originate from widespread societal perceptions on aging and technology, which commonly depict older adults as less skilled or less inclined to engage with new technology [76,77]. Such perspectives can unintentionally result in exclusion from IVR clinical trials or use in the clinical setting [78]. Thus, we may underestimate the older adults' capacity to accept or learn new technology such as IVR. Personalized education, tailored training, and sufficient support are essential actions to enhance the adoption of IVR among older adults [60].

Finally, aligning IVR content with older adults' personal interests, such as cultural and spiritual practices, as well as hobbies and previous life experiences, is vital for its wider acceptance in this demographic [30]. Future research is needed to investigate whether the level of pain or relaxation correlates with specific types of IVR content or exhibits a dose-response relationship based on duration and frequency of use. Finally, the insights gained from this study could pave the way for more extensive randomized controlled trials aimed at assessing the acceptance and overall efficacy of IVR in older adults undergoing a variety of major elective surgeries.

## Limitations

The findings of this study should be considered in view of certain limitations. The study design and limited sample size restrict the generalizability of the findings; therefore, final conclusions about the efficacy of IVR to improve pain and relaxation cannot be made. Additionally, this study was not designed or powered to reliably assess the impact of predictors such as age, gender, specific surgeries, or other baseline characteristics on study outcomes. Postoperative pain and relaxation outcomes are influenced by multiple perioperative factors that were not controlled in this initial feasibility study, including baseline pain levels prior to surgery, preoperative opioid consumption, extent of colorectal surgery, use of regional anesthesia or local blocks, and concurrent use of multimodal opioid-sparing medications. Moreover, the reliance on self-report introduces potential for bias and subjectivity in our findings. Although targeted recruitment efforts were undertaken to increase diversity, the study sample reflected the demographic composition of elective colorectal surgery patients treated within a single academic health care system during the study period. As a result, these findings may not fully capture the experiences of individuals from racially or ethnically diverse backgrounds, non-English speakers, those with lower educational attainment, or patients with frailty or cognitive impairment. This limitation highlights the importance of future multisite studies to more robustly evaluate the feasibility, accessibility, and broader applicability of IVR interventions in colorectal surgical populations. Also, because pain and relaxation outcomes were assessed only immediately before and after each IVR session, we were unable to determine the duration of the intervention's effects beyond the short-term period. Given the short average

postoperative length of stay in this cohort (just over 3 days), the intervention and assessments were necessarily limited to the immediate inpatient recovery window. It is key to note that although questionnaires were brief, repeated pre-post assessments across multiple IVR sessions may have contributed to respondent burden and fatigue in the immediate postoperative setting, although not mentioned by participants. Survey methodology literature suggests that repeated measures can affect response quality. Importantly, the assessment schedule remained feasible in this pilot study; however, larger and more longitudinal studies should explicitly evaluate participant burden related to repeated questionnaires and assessment frequency [79]. Finally, the IVR intervention was not standardized, allowing variation in both content selection and duration of use. Given that some participants selected briefer immersive content experiences while others completed longer sessions, it remains unclear whether shorter exposure reflected content preference, postoperative fatigue, discomfort, or other barriers to sustained use. Future studies should consider controlling content type, duration, and frequency of use to better assess the impact of IVR on acceptability, feasibility, and clinical outcomes.

## Conclusion

This study supports the feasibility, acceptability, and tolerability of IVR as a potential intervention for postoperative pain among older adults following elective inpatient abdominal surgery. Our findings add to the growing body of evidence supporting nonpharmacological approaches for postoperative pain management. While our preliminary findings are promising, larger-scale studies are needed to confirm the acceptance and efficacy of IVR as a postoperative pain management intervention across more diverse populations of older adults, including those from underrepresented minority groups and individuals with physical or cognitive limitations. Importantly, this study highlights the potential of IVR to enhance patient-reported outcomes and improve the perioperative care experience for a demographic that is often considered vulnerable and is frequently underrepresented in technology-based research.

## Acknowledgments

We declare that artificial intelligence was not used in any aspect of this research study or manuscript development.

## Funding

This work was supported by the National Institute of Nursing Research of the National Institutes of Health (T32NR01692), the University of California, San Francisco (UCSF), School of Nursing Century Research award, the UCSF Graduate Mentorship Fellowship award, the UCSF Phillip R. Lee Health Policy Fellowship award, the UCSF Office of the Executive Vice Chancellor and Provost (EVCP) Social and Population Science Fellowship Research Award, and by the Department of Veterans Affairs, Veterans Health Administration, Office of Academic Affiliations VA Quality Scholars Advanced Fellowship Program (OAA 3Q072019C).

## Conflicts of Interest

None declared.

## References

1. Schwarze ML, Barnato AE, Rathouz PJ, et al. Development of a list of high-risk operations for patients 65 years and older. JAMA Surg 2015 Apr;150(4):325-331. [doi: 10.1001/jamasurg.2014.1819] [Medline: 25692282]
2. Cheung C, Meissner MA, Garg T. Incorporating outcomes that matter to older adults into surgical research. J Am Geriatr Soc 2021 Mar;69(3):618-620. [doi: 10.1111/jgs.17028] [Medline: 33462830]
3. Aubrun F, Marmion F. The elderly patient and postoperative pain treatment. Best Pract Res Clin Anaesthesiol 2007 Mar;21(1):109-127. [doi: 10.1016/j.bpa.2006.12.005] [Medline: 17489223]
4. Hwang SS, Chang VT, Kasimis B. Dynamic cancer pain management outcomes: the relationship between pain severity, pain relief, functional interference, satisfaction and global quality of life over time. J Pain Symptom Manage 2002 Mar;23(3):190-200. [doi: 10.1016/s0885-3924(01)00418-3] [Medline: 11888717]
5. Apfelbaum JL, Chen C, Mehta SS, Gan TJ. Postoperative pain experience: results from a national survey suggest postoperative pain continues to be undermanaged. Anesth Analg 2003 Aug;97(2):534-540. [doi: 10.1213/01.ANE.0000068822.10113.9E] [Medline: 12873949]
6. Borys M, Zyzak K, Hanych A, et al. Survey of postoperative pain control in different types of hospitals: a multicenter observational study. BMC Anesthesiol 2018 Jul 18;18(1):83. [doi: 10.1186/s12871-018-0551-3] [Medline: 30021520]
7. Jin F, Chung F. Minimizing perioperative adverse events in the elderly. Br J Anaesth 2001 Oct;87(4):608-624. [doi: 10.1093/bja/87.4.608] [Medline: 11878732]
8. Falzone E, Hoffmann C, Keita H. Postoperative analgesia in elderly patients. Drugs Aging 2013 Feb;30(2):81-90. [doi: 10.1007/s40266-012-0047-7] [Medline: 23288604]
9. Kessler ER, Shah M, Gruschkus SK, Raju A. Cost and quality implications of opioid-based postsurgical pain control using administrative claims data from a large health system: opioid-related adverse events and their impact on clinical and economic outcomes. Pharmacotherapy 2013 Apr;33(4):383-391. [doi: 10.1002/phar.1223] [Medline: 23553809]

10. Baker J, Brovman EY, Rao N, Beutler SS, Urman RD. Potential opioid-related adverse drug events are associated with decreased revenue in hip replacement surgery in the older population. Geriatr Orthop Surg Rehabil 2020;11:2151459320915328. [doi: 10.1177/2151459320915328] [Medline: 32231864]
11. Huhn AS, Strain EC, Tompkins DA, Dunn KE. A hidden aspect of the U.S. opioid crisis: rise in first-time treatment admissions for older adults with opioid use disorder. Drug Alcohol Depend 2018 Dec 1;193:142-147. [doi: 10.1016/j.drugalcdep.2018.10.002] [Medline: 30384321]
12. Hamilton HJ, Gallagher PF, O'Mahony D. Inappropriate prescribing and adverse drug events in older people. BMC Geriatr 2009 Jan 28;9:5. [doi: 10.1186/1471-2318-9-5] [Medline: 19175914]
13. Mangoni AA, Jackson SHD. Age-related changes in pharmacokinetics and pharmacodynamics: basic principles and practical applications. Br J Clin Pharmacol 2004 Jan;57(1):6-14. [doi: 10.1046/j.1365-2125.2003.02007.x] [Medline: 14678335]
14. McKeown JL. Pain management issues for the geriatric surgical patient. Anesthesiol Clin 2015 Sep;33(3):563-576. [doi: 10.1016/j.anclin.2015.05.010] [Medline: 26315638]
15. Weiss AJ, Heslin KC, Barrett ML, Izar R, Bierman AS. Opioid-Related Inpatient Stays and Emergency Department Visits Among Patients Aged 65 Years and Older, 2010 and 2015: Agency for Healthcare Research and Quality (US); 2018. URL: https://www.ncbi.nlm.nih.gov/books/NBK534106 [accessed 2026-03-17]
16. Non-pharmacologic and non-opioid solutions for pain management. : The Joint Commission; 2018 URL: https://digitalassets.jointcommission.org/api/public/content/b79588e914734c62ba3d7a4427ee3a5a?v=84bb0b4d [accessed 2026-04-09]
17. Martinez V, Beloeil H, Marret E, Fletcher D, Ravaud P, Trinquart L. Non-opioid analgesics in adults after major surgery: systematic review with network meta-analysis of randomized trials. Br J Anaesth 2017 Jan;118(1):22-31. [doi: 10.1093/bja/aew391] [Medline: 28039239]
18. Opioids in medicare part D: concerns about extreme use and questionable prescribing. : U.S. Department of Health & Human Services; 2017 URL: https://oig.hhs.gov/OEI/REPORTS/OEI-02-17-00250.pdf [accessed 2026-03-17]
19. Baker NA, Polhemus AH, Haan Ospina E, et al. The state of science in the use of virtual reality in the treatment of acute and chronic pain: a systematic scoping review. Clin J Pain 2022 Jun 1;38(6):424-441. [doi: 10.1097/AJP.0000000000001029] [Medline: 35537072]
20. Furht B, editor. Immersive virtual reality. In: Encyclopedia of Multimedia: Springer; 2008:345-346. [doi: 10.1007/978-0-387-78414-4_85]
21. Gupta A, Scott K, Dukewich M. Innovative technology using virtual reality in the treatment of pain: does It reduce pain via distraction, or is there more to it? Pain Med 2018 Jan 1;19(1):151-159. [doi: 10.1093/pm/pnx109] [Medline: 29025113]
22. Malloy KM, Milling LS. The effectiveness of virtual reality distraction for pain reduction: a systematic review. Clin Psychol Rev 2010 Dec;30(8):1011-1018. [doi: 10.1016/j.cpr.2010.07.001] [Medline: 20691523]
23. Dan A, Reiner M. EEG-based cognitive load of processing events in 3D virtual worlds is lower than processing events in 2D displays. Int J Psychophysiol 2017 Dec;122:75-84. [doi: 10.1016/j.ijpsycho.2016.08.013] [Medline: 27592084]
24. Chan E, Foster S, Sambell R, Leong P. Clinical efficacy of virtual reality for acute procedural pain management: a systematic review and meta-analysis. PLoS One 2018;13(7):e0200987. [doi: 10.1371/journal.pone.0200987] [Medline: 30052655]
25. Ding L, Hua H, Zhu H, et al. Effects of virtual reality on relieving postoperative pain in surgical patients: a systematic review and meta-analysis. Int J Surg 2020 Oct;82:87-94. [doi: 10.1016/j.ijsu.2020.08.033] [Medline: 32882400]
26. Iannicelli AM, Vito D, Dodaro CA, et al. Does virtual reality reduce pain in pediatric patients? A systematic review. Ital J Pediatr 2019 Dec 30;45(1):171. [doi: 10.1186/s13052-019-0757-0] [Medline: 31888710]
27. Syed-Abdul S, Malwade S, Nursetyo AA, et al. Virtual reality among the elderly: a usefulness and acceptance study from Taiwan. BMC Geriatr 2019 Aug 19;19(1):223. [doi: 10.1186/s12877-019-1218-8] [Medline: 31426766]
28. Seifert A, Schlomann A. The use of virtual and augmented reality by older adults: potentials and challenges. Front Virtual Real 2021;2. [doi: 10.3389/frvir.2021.639718]
29. Healy D, Carr E, Conlan O, Browne AC, Walsh JC. Exploring the content of the STAND-VR intervention: a qualitative interview study. PLoS Digit Health 2023 Mar;2(3):e0000210. [doi: 10.1371/journal.pdig.0000210] [Medline: 36913343]
30. Healy D, Flynn A, Conlan O, McSharry J, Walsh J. Older adults' experiences and perceptions of immersive virtual reality: systematic review and thematic synthesis. JMIR Serious Games 2022 Dec 6;10(4):e35802. [doi: 10.2196/35802] [Medline: 36472894]
31. Benham S, Kang M, Grampurohit N. Immersive virtual reality for the management of pain in community-dwelling older adults. OTJR (Thorofare N J) 2019 Apr;39(2):90-96. [doi: 10.1177/1539449218817291] [Medline: 30595096]
32. Goudman L, Jansen J, Billot M, et al. Virtual reality applications in chronic pain management: systematic review and meta-analysis. JMIR Serious Games 2022 May 10;10(2):e34402. [doi: 10.2196/34402] [Medline: 35536641]
33. Peng L, Zeng Y, Wu Y, Si H, Shen B. Virtual reality-based rehabilitation in patients following total knee arthroplasty: a systematic review and meta-analysis of randomized controlled trials. Chin Med J (Engl) 2021 Dec 13;135(2):153-163. [doi: 10.1097/CM9.0000000000001847] [Medline: 34908004]
34. Su W, Zhou Y, Qiu H, Wu H. The effects of preoperative rehabilitation on pain and functional outcome after total knee arthroplasty: a meta-analysis of randomized controlled trials. J Orthop Surg Res 2022 Mar 21;17(1):175. [doi: 10.1186/s13018-022-03066-9] [Medline: 35313897]

35. Huang Q, Lin J, Han R, Peng C, Huang A. Using virtual reality exposure therapy in pain management: a systematic review and meta-analysis of randomized controlled trials. Value Health 2022 Feb;25(2):288-301. [doi: 10.1016/j.jval.2021.04.1285] [Medline: 35094802]
36. Eldridge SM, Chan CL, Campbell MJ, et al. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. Pilot Feasibility Stud 2016;2(1):64. [doi: 10.1186/s40814-016-0105-8] [Medline: 27965879]
37. Lancaster GA, Thabane L. Guidelines for reporting non-randomised pilot and feasibility studies. Pilot Feasibility Stud 2019;5(1):114. [doi: 10.1186/s40814-019-0499-1] [Medline: 31608150]
38. Birckhead B, Khalil C, Liu X, et al. Recommendations for methodology of virtual reality clinical trials in health care by an International Working Group: iterative study. JMIR Ment Health 2019 Jan 31;6(1):e11973. [doi: 10.2196/11973] [Medline: 30702436]
39. Immersive healthcare. Penumbra. URL: https://www.penumbrainc.com/immersive-healthcare/ [accessed 2026-04-09]
40. Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG. Simulator Sickness Questionnaire: an enhanced method for quantifying simulator sickness. Int J Aviat Psychol 1993 Jul;3(3):203-220. [doi: 10.1207/s15327108ijap0303_3]
41. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Stud 2009;4(3):114-123 [FREE Full text]
42. Brooke J. SUS: a “quick and dirty” usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland IL, editors. Usability Evaluation in Industry: Taylor & Francis; 1996:207-212. [doi: 10.1201/9781498710411-35]
43. Brooks R. EuroQol: the current state of play. Health Policy 1996 Jul;37(1):53-72. [doi: 10.1016/0168-8510(96)00822-6] [Medline: 10158943]
44. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res 2011 Dec;20(10):1727-1736. [doi: 10.1007/s11136-011-9903-x] [Medline: 21479777]
45. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing Generalized Anxiety Disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
46. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613. [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
47. Sullivan MJL, Bishop SR, Pivik J. The Pain Catastrophizing Scale: development and validation. Psychol Assess 1995;7(4):524-532. [doi: 10.1037/1040-3590.7.4.524]
48. Sullivan MJL. The complete statistical software for data science. Stata. URL: https://www.stata.com/ [accessed 2021-03-10]
49. Schrempf MC, Zanker J, Arndt TT, et al. Immersive virtual reality fitness games to improve recovery after colorectal surgery: a randomized single blind controlled pilot trial. Games Health J 2023 Dec;12(6):450-458. [doi: 10.1089/g4h.2023.0004] [Medline: 37428543]
50. Ugras GA, Kanat C, Yaman Z, Yilmaz M, Turkmenoglu MO. The effects of virtual reality on preoperative anxiety in patients undergoing colorectal and abdominal wall surgery: a randomized controlled trial. J Perianesth Nurs 2023 Apr;38(2):277-283. [doi: 10.1016/j.jopan.2022.07.005] [Medline: 36319521]
51. Smith RC. Making the biopsychosocial model more scientific-its general and specific models. Soc Sci Med 2021 Mar;272:113568. [doi: 10.1016/j.socscimed.2020.113568] [Medline: 33423810]
52. Wang S, Lim SH, Aloweni F. Virtual reality interventions and the outcome measures of adult patients in acute care settings undergoing surgical procedures: an integrative review. J Adv Nurs 2022 Mar;78(3):645-665. [doi: 10.1111/jan.15065] [Medline: 34633112]
53. Tashjian VC, Mosadeghi S, Howard AR, et al. Virtual reality for management of pain in hospitalized patients: results of a controlled trial. JMIR Ment Health 2017 Mar 29;4(1):e9. [doi: 10.2196/mental.7387] [Medline: 28356241]
54. Mosadeghi S, Reid MW, Martinez B, Rosen BT, Spiegel BMR. Feasibility of an immersive virtual reality intervention for hospitalized patients: an observational cohort study. JMIR Ment Health 2016 Jun 27;3(2):e28. [doi: 10.2196/mental.5801] [Medline: 27349654]
55. Hauk N, Hüffmeier J, Krumm S. Ready to be a silver surfer? A meta-analysis on the relationship between chronological age and technology acceptance. Comput Human Behav 2018 Jul;84:304-319. [doi: 10.1016/j.chb.2018.01.020]
56. Dermody G, Whitehead L, Wilson G, Glass C. The role of virtual reality in improving health outcomes for community-dwelling older adults: systematic review. J Med Internet Res 2020 Jun 1;22(6):e17331. [doi: 10.2196/17331] [Medline: 32478662]
57. Moore RC, Hancock JT, Bailenson JN. From 65 to 103, older adults experience virtual reality differently depending on their age: evidence from a large-scale field study in nursing homes and assisted living facilities. Cyberpsychol Behav Soc Netw 2023 Dec;26(12):886-895. [doi: 10.1089/cyber.2023.0188] [Medline: 38011717]
58. Kelleher SA, Fisher HM, Winger JG, et al. Virtual reality for improving pain and pain-related symptoms in patients with advanced stage colorectal cancer: a pilot trial to test feasibility and acceptability. Palliat Support Care 2022 Aug;20(4):471-481. [doi: 10.1017/S1478951521002017] [Medline: 35078545]
59. Hawley-Hague H, Boulton E, Hall A, Pfeiffer K, Todd C. Older adults’ perceptions of technologies aimed at falls prevention, detection or monitoring: a systematic review. Int J Med Inform 2014 Jun;83(6):416-426. [doi: 10.1016/j.ijmedinf.2014.03.002] [Medline: 24798946]

60. Roberts AR, De Schutter B, Franks K, Radina ME. Older adults' experiences with audiovisual virtual reality: perceived usefulness and other factors influencing technology acceptance. Clin Gerontol 2019;42(1):27-33. [doi: 10.1080/07317115.2018.1442380] [Medline: 29505343]
61. Mitzner TL, Boron JB, Fausset CB, et al. Older adults talk technology: technology usage and attitudes. Comput Human Behav 2010 Nov 1;26(6):1710-1721. [doi: 10.1016/j.chb.2010.06.020] [Medline: 20967133]
62. Lee LN, Kim MJ, Hwang WJ. Potential of augmented reality and virtual reality technologies to promote wellbeing in older adults. Appl Sci (Basel) 2019;9(17):3556. [doi: 10.3390/app9173556]
63. Rawlins CR, Veigulis Z, Hebert C, Curtin C, Osborne TF. Effect of immersive virtual reality on pain and anxiety at a Veterans Affairs health care facility. Front Virtual Real 2021;2. [doi: 10.3389/frvir.2021.719681]
64. Bekelis K, Calnan D, Simmons N, MacKenzie TA, Kakoulides G. Effect of an immersive preoperative virtual reality experience on patient reported outcomes: a randomized controlled trial. Ann Surg 2017 Jun;265(6):1068-1073. [doi: 10.1097/SLA.0000000000002094] [Medline: 27906757]
65. Huang MY, Scharf S, Chan PY. Effects of immersive virtual reality therapy on intravenous patient-controlled sedation during orthopaedic surgery under regional anesthesia: a randomized controlled trial. PLoS One 2020;15(2):e0229320. [doi: 10.1371/journal.pone.0229320] [Medline: 32092098]
66. Tian N, Lopes P, Boulic R. A review of cybersickness in head-mounted displays: raising attention to individual susceptibility. Virtual Real 2022 Dec;26(4):1409-1441. [doi: 10.1007/s10055-022-00638-2]
67. Drazich BF, McPherson R, Gorman EF, et al. In too deep? A systematic literature review of fully-immersive virtual reality and cybersickness among older adults. J Am Geriatr Soc 2023 Dec;71(12):3906-3915. [doi: 10.1111/jgs.18553] [Medline: 37560978]
68. Facebook technologies recalls removable foam facial interfaces for oculus quest 2 virtual reality headsets due to skin irritation hazard (recall alert). US Consumer Product Safety Commission. URL: https://www.cpsc.gov/Recalls/2021/Facebook-Technologies-Recalls-Removable-Foam-Facial-Interfaces-for-Oculus-Quest-2-Virtual-Reality-Headsets-Due-to-Skin-Irritation-Hazard-Recall-Alert [accessed 2024-04-01]
69. Høeg ER, Andersen NB, Malmkjær N, Vaaben AH, Uth J. Hospitalized older adults' experiences of virtual reality-based group exercise therapy with cycle ergometers: an early feasibility study. Comput Hum Behav Rep 2023 Aug;11:100301. [doi: 10.1016/j.chbr.2023.100301]
70. Delshad SD, Almario CV, Fuller G, Luong D, Spiegel BMR. Economic analysis of implementing virtual reality therapy for pain among hospitalized patients. NPJ Digital Med 2018;1(1):1-8. [doi: 10.1038/s41746-018-0026-4]
71. 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. [doi: 10.1097/MLR.0000000000000625] [Medline: 27623005]
72. Pandrangi VC, Shah SN, Bruening JD, et al. Effect of virtual reality on pain management and opioid use among hospitalized patients after head and neck surgery: a randomized clinical trial. JAMA Otolaryngol Head Neck Surg 2022 Aug 1;148(8):724-730. [doi: 10.1001/jamaoto.2022.1121] [Medline: 35679057]
73. Slomski A. Virtual reality lessens postoperative pain. JAMA 2022 Aug 16;328(7):610. [doi: 10.1001/jama.2022.13772] [Medline: 35972497]
74. Geriatric surgery verification. American College of Surgeons. URL: https://www.facs.org/quality-programs/accreditation-and-verification/geriatric-surgery-verification/ [accessed 2026-04-09]
75. CMS issues first HCPCS code and Medicare DME benefit category determination for therapeutic Virtual Reality Device. DLA Piper. URL: https://www.dlapiper.com/en-ca/insights/publications/2023/04/cms-issues-first-hcpcs-code-and-medicare-dme-benefit-category-determination [accessed 2024-02-24]
76. Heinz M, Martin P, Margrett JA, et al. Perceptions of technology among older adults. J Gerontol Nurs 2013 Jan;39(1):42-51. [doi: 10.3928/00989134-20121204-04] [Medline: 23244061]
77. Vaportzis E, Clausen MG, Gow AJ. Older adults perceptions of technology and barriers to interacting with tablet computers: a focus group study. Front Psychol 2017 Oct 4;8:1687. [doi: 10.3389/fpsyg.2017.01687] [Medline: 29071004]
78. Mace RA, Mattos MK, Vranceanu AM. Older adults can use technology: why healthcare professionals must overcome ageism in digital health. Transl Behav Med 2022 Dec 30;12(12):1102-1105. [doi: 10.1093/tbm/ibac070] [Medline: 36073770]
79. Rolstad S, Adler J, Rydén A. Response burden and questionnaire length: is shorter better? A review and meta-analysis. Value Health 2011 Dec;14(8):1101-1108. [doi: 10.1016/j.jval.2011.06.003] [Medline: 22152180]


## Abbreviations

**CONSORT:** Consolidated Standards of Reporting Trials
**GAD-7:** Generalized Anxiety Disorder 7-item scale
**HMD:** head-mounted display
**IVR:** immersive virtual reality
**NRS:** Numeric Rating Scale
**PCS:** Pain Catastrophizing Scale
**PHQ-8:** Patient Health Questionnaire-8

**SPMSQ:** Short Portable Mental Status Questionnaire
**SSQ:** Simulator Sickness Questionnaire
**SUS:** System Usability Scale
**VAS:** Visual Analog Scale
**VR-CORE:** Virtual Reality Clinical Outcomes Research Experts

*Edited by J Shiffermiller; submitted 03.Aug.2025; peer-reviewed by S Langenfeld, Sunny, CL Au; revised version received 28.Feb.2026; accepted 03.Mar.2026; published 17.Apr.2026.*

*Please cite as:*
*Keny C, Shafiq U, Lorenz K, Russell M, Leutwyler H, Wagner LM, Tang V, Park LG*
*Immersive Virtual Reality for Pain and Relaxation in Older Adults Following Elective Inpatient Abdominal Surgery: Single-Arm Study Examining Feasibility and Acceptability*
*JMIR Perioper Med 2026;9:e81791*
*URL: https://periop.jmir.org/2026/1/e81791*
*doi:10.2196/81791*

# Association Between Complications and Death Within 30 Days After Orthopedic Surgery: Vascular Events in Noncardiac Surgery Patients Cohort Evaluation (VISION) Substudy

Lily J Park[1,2], MD, MSc; P J Devereaux[2,3,4], MD, PhD; Ameen Patel[2,5], MD; Vikas Tandon[2,5], MD; Diane Heels-Ansdell[6], MSc; Lehana Thabane[6], PhD; Pablo E Serrano[1], MD, PhD; Matthew T V Chan[7], MBBS, PhD; Wojciech Szczeklik[8], MD, PhD; Sadeesh Srinathan[9], MD, MSc; Ignacio Garutti[10], MD, PhD; Gerard Urrutia[11], MS, MD, PhD; Ernesto Guerra-Farfan[12], MD; Hassaan Abdel Khalik[13], MD; Emmanuelle Duceppe[14], MD, PhD; Sandra Ofori[2,3], MSc, MD, MBBS, PhD; Maura Marcucci[2], MD, MSc; David Conen[2,3], MD, MPH; Michael K Wang[2,3], MD; Jessica Spence[2,15], MD, PhD; Daniel Tushinski[13], MD, MSc; Kamal Bali[13], MBBS, MSc; Anthony Adili[13], MD; Vickas Khanna[13], MD, MHA; Ana Claudia Tonelli[16], MD, PhD; Francesca Mulazzani[17], MD; Wenjun Jiang[18]; Olufemi R Ayeni[13], MD, PhD; Gerard Slobegean[19], MD, MPH; Theodore Miclau[20], MD; Mohit Bhandari[6,13], MSc, MD, PhD; Flavia K Borges[2,5,6], MD, PhD

[1]Department of Surgery, Division of General Surgery, McMaster University, Hamilton, ON, Canada
[10]Department of Medicine, Universidad Complutense de Madrid, Madrid, Spain
[11]Institut de Recerca Sant Pau (IR SANT PAU) – CIBERESP, Barcelona, Spain
[12]Department of Surgery, Division of Orthopedic surgery, Vall d'Hebron Hospital Universitari, Barcelona, Catalonia, Spain
[13]Department of Surgery, Division of Orthopedic Surgery, McMaster University, Hamilton, ON, Canada
[14]Department of Medicine, Université de Montréal, Montreal, QC, Canada
[15]Department of Anesthesia, McMaster University, Hamilton, ON, Canada
[16]Internal Medicine Service, Hospital de Clínicas de Porto Alegre, Porto Alegre, Rio Grande do Sul, Brazil
[17]Department of Medicine, University of Milano-Bicocca, Milan, Lombardy, Italy
[18]Faculty of Health Sciences, McMaster University, Hamilton, ON, Canada
[19]Department of Orthopedics, School of Medicine, University of California, Irvine, CA, United States
[2]Population Health Research Institute, 237 Barton St E, C1-109, Hamilton, ON, Canada
[20]Department of Orthopaedic Surgery, University of California, San Francisco, San Francisco, CA, United States
[3]Division of Cardiology, Department of Medicine, McMaster University, Hamilton, ON, Canada
[4]World Health Research, Hamilton, ON, Canada
[5]Department of Medicine, McMaster University, Hamilton, ON, Canada
[6]Department of Health Research Methods, Evidence & Impact, McMaster University, Hamilton, ON, Canada
[7]Chinese University of Hong Kong, Hong Kong, China (Hong Kong)
[8]Centre for Intensive Care and Perioperative Medicine, Jagiellonian College, Krakow, Kujawsko-Pomorskie, Poland
[9]Department of Surgery, University of Manitoba, Winnipeg, MB, Canada

**Corresponding Author:**
Flavia K Borges, MD, PhD
Population Health Research Institute, 237 Barton St E, C1-109, Hamilton, ON, Canada

## Abstract

**Background:** The contemporary causes of postoperative mortality in orthopedic surgery are not well characterized.

**Objective:** This study aimed to describe the epidemiology of postoperative complications among adult patients who underwent orthopedic surgery and inform their relationships with 30-day mortality.

**Methods:** Vascular Events in Noncardiac Surgery Patients Cohort Evaluation (VISION) was a prospective cohort study involving 40,004 adult patients who underwent noncardiac surgery across 28 centers in 14 countries. For the subset of patients who underwent orthopedic surgery, a Cox proportional hazards model was used to determine time-dependent associations between various surgical complications and 30-day postoperative mortality. Analyses were adjusted for preoperative and surgical variables.

**Results:** Among 8385 patients who underwent an orthopedic surgery in VISION, 1.6% (n=132) died within 30 days of surgery. Of the 132 deaths, 63.6% (n=84) occurred in hospital during the index hospitalization, while 36.4% (n=48) occurred after discharge.

The incidence of death across the subcategories of orthopedic surgery was above-knee amputation (30/221, 13.6%), internal fixation of femur (29/750, 3.9%), lower leg amputation (9/252, 3.6%), major hip or pelvic surgery (49/2898, 1.7%), major spine surgery (8/1405, 0.6%), and knee arthroplasty (7/2876, 0.2%). A total of 6 postoperative complications (myocardial injury after noncardiac surgery [MINS], major bleeding, infection without sepsis, sepsis, stroke, and atrial fibrillation) were associated with death in adjusted analyses. The greatest attributable fractions of postoperative mortality (ie, proportion of deaths in the cohort that can be attributed to each complication, if causality were established) were from MINS (1454/8385, 17.3%; hazard ratio [HR] 2.08, 95% CI 1.38 - 3.14; *P*<.001; attributable fraction=20.6%), major bleeding (2422/8385, 28.9%; HR 1.95, 95% CI 1.34 - 2.85; *P*<.001; attributable fraction=16.5%), and sepsis (318/8385, 3.8%; HR 6.24, 95% CI 3.85 - 10.12; *P*<.001; attributable fraction=9.7%).

**Conclusions:** The complications most attributable to 30-day mortality following orthopedic surgery were MINS, major bleeding, and sepsis. These findings highlight areas for further study to mitigate perioperative mortality in orthopedic surgery. MINS demonstrated the highest attributable fraction for mortality (20.6%), emphasizing the importance of appropriate MINS screening, diagnosis, and management.

***(JMIR Perioper Med 2026;9:e90823)*** doi:10.2196/90823

**KEYWORDS**

perioperative medicine; orthopedic surgery; morbidity; postoperative complications; mortality

## Introduction

Globally, the number of orthopedic surgeries performed annually is estimated to have approached 31.4 million procedures in 2024 [1]. Among these patients, the mortality rate is broadly estimated to be between 0.6% to 8.7%. Even at the lower end of this estimated range, mortality after orthopedic surgery represents a substantial health issue [2-5]. Much of the existing mortality data are derived from patients with hip fracture and orthopedic trauma, which represents only a small fraction of the entire orthopedic surgery population [2-5]. This highlights the need to better elucidate mortality rates in a broad sample of general contemporary patients who underwent orthopedic surgery.

The largest study to date in this area used the National Hospital Discharge Survey to identify risk factors for mortality in orthopedic surgery across a nationwide sample of hospitals in the United States [5]. However, this study is limited by the retrospective nature of the study, reliance on administrative data, and older data. Furthermore, the National Hospital Discharge Survey is considered less reliable in reflecting morbidity and complications [5-10]. Considering this, an updated and accurate understanding of modifiable risk factors for death in orthopedic surgery is needed.

Vascular Events in Noncardiac Surgery Patients Cohort Evaluation (VISION) was a large prospective cohort study that included patients who underwent noncardiac surgery and were systematically followed to document postoperative complications, including mortality [11]. We previously reported the incidence of perioperative complications and associated mortality for the entire VISION study population. Patients undergoing orthopedic surgery represent a unique population that increasingly includes older adults with rising prevalence of cardiovascular disease [12-14]. This necessitates specialty-specific epidemiologic data.

The objective of this prospective cohort substudy was to describe the epidemiology of postoperative complications and death, evaluate the associations between these postoperative complications and death, and report the attributable fractions of each complication for death among a contemporary orthopedic surgery cohort.

## Methods

### Ethical Considerations

The research ethics board at each participating site approved the protocol before patient recruitment. This study was approved by the Hamilton Integrated Research Ethics Board under the project number 07-220. All participants provided informed consent and we followed the ethical standard principles of the Declaration of Helsinki, including maintaining privacy and confidentiality of research participants' data. No compensation was provided to participants.

### Study Design

The design and methods of the VISION study have been previously described [11,15]. In summary, VISION was an international prospective cohort study, which enrolled 40,004 patients who underwent noncardiac surgery across 28 centers in 14 countries from August 2007 to November 2013. Patients aged ≥45 years, who underwent noncardiac surgery, receiving general or regional anesthesia, and requiring at least 1 overnight hospital stay after surgery were eligible for inclusion. This inclusion criteria was determined a priori with the intention to efficiently identify those who are at higher risk of mortality and capture a greater number of mortality outcomes and perioperative complications. Each participating hospital obtained approval from their research ethics board prior to the start of patient enrollment. Refer to Multimedia Appendix 1 for funding sources. The Strengthening the Reporting of Observational Studies in Epidemiology Statement was followed (Checklist 1).

### Analysis Population

Patients who were enrolled in the VISION study and underwent an orthopedic procedure (including spine surgery) were included in this substudy. According to the VISION definitions, this would include surgeries categorized as major hip or pelvic surgery, internal fixation of femur, knee arthroplasty,

above-knee amputation, lower leg amputation, and major spine surgery.

### Follow-Up

Patients were followed for 30 days following their surgery and were censored at the time of their last assessment if the 30-day follow-up was not complete.

### Complications Variables

Definitions for all variables included in the analyses are provided in (Multimedia Appendix 2). The primary outcome was time to all-cause mortality. The following postoperative complications during the first 30 days after surgery were investigated: myocardial injury after noncardiac surgery (MINS), venous thromboembolism (VTE), stroke, major bleeding, acute kidney injury (AKI) resulting in dialysis, sepsis, infection without sepsis, new clinically important atrial fibrillation, and congestive heart failure [11,12,16-19]. These variables were selected for clinical relevance while being parsimonious to preserve model stability. Of note, major bleeding was defined using a previously validated definition in the perioperative setting demonstrating independent association with 30-day mortality [18,19].

### Statistical Analysis

A statistical analysis plan was written before undertaking analyses for this study, which was finalized in June 2023. Descriptive statistics were used to report patient characteristics, as well as the incidence of death and postoperative complications. As AKI resulting in dialysis occurred at the rate of 0.2%, this variable was omitted from the model to preserve model stability.

To determine the relationship between complications and 30-day mortality, we used a time-dependent Cox proportional hazards model and adjusted for baseline characteristics that were known to be independently associated with 30-day mortality, according to previous VISION analyses [11]. Complications were modeled as time-dependent covariates, allowing for multiple events per patient and accounting for their timing relative to death. This approach enabled assessment of the association between each complication with mortality rather than assignment of a single causal event. This model was adjusted for age category (65 - 75 vs 45 - 65 years, and age ≥75 vs 45 - 65 years), cancer at the time of surgery, history of chronic obstructive pulmonary disease, surgery urgency (emergent: <24 hours from admission for an acute surgical condition, urgent: 24 - 72 hours from admission for an acute surgical condition, nonemergent: all other surgeries), and history of peripheral arterial disease. All variables included in the model were selected a priori. We reported the adjusted hazard ratios (HRs) and corresponding 95% CIs. To avoid optimistic estimation of the c-statistic from possible overfitting, model performance was assessed using the c-statistic corrected for optimism using 1000 bootstrapped samples [20]. This involved the generation of 1000 simulated datasets through random sampling and replacement from the original dataset to estimate model performance while accounting for overfitting. Attributable fractions for death were calculated using an established method [21]. The attributable fraction represents the proportion of deaths that would not have occurred in the VISION orthopedic cohort if the complication had not occurred, if we assumed a causal relationship between the complication and death. Attributable fraction methods have been increasingly used in perioperative and epidemiologic research. Prior applications include large cohort analyses in noncardiac surgery and population health studies evaluating modifiable contributors to mortality [11,17,21-23]. There were 10 events per variable included in the model, which supports model stability [24].

For all tests, we used $\alpha$ <.05 as the level of significance. Analyses were performed using R statistical software (version 4.2.2; R Foundation for Statistical Computing) using the "survival," "epiR," and "ggplot2" packages.

## Results

### Patient Characteristics

Within the 40,004 patients enrolled in the overall VISION study, 21% (8385/40,004) underwent an orthopedic surgery and were included in these analyses. Figure 1 demonstrates the flow of patient inclusion. Table 1 presents the baseline characteristics of the participants. More than half of the patients were aged >64 years, and 57.3% (4802/8385) were female. There was no loss to follow-up in the orthopedic surgery population.

**Figure 1.** Flow of patient inclusion in the Vascular Events in Noncardiac Surgery Patients Cohort Evaluation (VISION) orthopedic surgery substudy. Note: in the overall VISION study, 284 patients had 2 or 3 surgical categories.

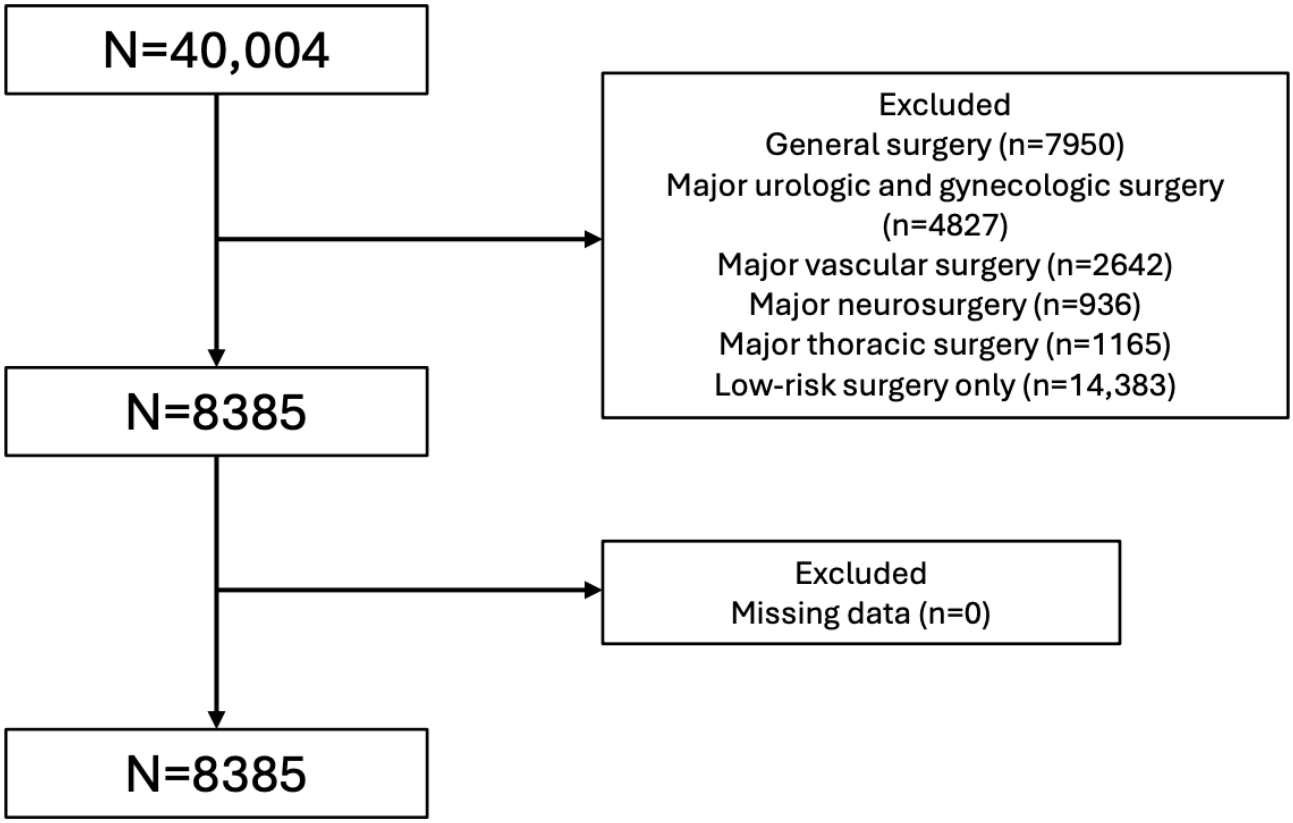


**Table .** Baseline characteristics of the orthopedic surgery cohort.

| Characteristics | Participants, n (%) | Number of deaths within 30 days, n (%) |
|---|---|---|
| Age (years) | | |
| 45 - 64 | 3411 (40.7) | 31 (0.9) |
| 65 - 74 | 2418 (28.8) | 20 (0.8) |
| ≥75 | 2556 (30.5) | 81 (3.2) |
| Sex: female | 4802 (57.3) | 76 (1.6) |
| History of | | |
| Hypertension | 4948 (59) | 89 (1.8) |
| Diabetes | 1823 (21.8) | 44 (2.4) |
| Coronary artery disease | 1253 (15) | 41 (3.3) |
| Peripheral arterial disease | 519 (6.2) | 39 (7.5) |
| Chronic obstructive pulmonary disease | 640 (7.6) | 33 (5.2) |
| Coronary revascularization | 470 (5.6) | 6 (1.3) |
| Stroke | 398 (4.7) | 20 (5) |
| Congestive heart failure | 401 (4.8) | 20 (5) |
| Active cancer | 320 (3.8) | 13 (4.1) |
| Atrial fibrillation | 323 (3.9) | 11 (3.4) |
| Preoperative estimated glomerular filtration rate (mL/min/1.73 $m^2$) | | |
| <30 | 294 (3.7) | 26 (8.8) |
| 30 - 44 | 503 (6.2) | 22 (4.4) |
| 45 - 59 | 988 (12.3) | 12 (1.2) |
| ≥60 | 6264 (77.8) | 67 (1.1) |
| Types of orthopedic surgery | | |
| Major hip or pelvic surgery | 2898 (34.6) | 49 (1.7) |
| Internal fixation of femur | 750 (8.9) | 29 (3.9) |
| Knee arthroplasty | 2876 (34.3) | 7 (0.2) |
| Above-knee amputation | 221 (2.6) | 30 (13.6) |
| Lower leg amputation | 252 (3) | 9 (3.6) |
| Major spine surgery | 1405 (16.8) | 8 (0.6) |
| Subcategory of surgery | | |
| Nonemergent | 7369 (87.9) | 81 (1.1) |
| Urgent or emergent | 1016 (12.1) | 51 (5) |
| Type of anesthesia | | |
| General only | 2436 (29.1) | 36 (1.5) |
| Neuraxial (spinal or epidural) only | 4182 (49.9) | 72 (1.7) |
| General with nitrous oxide only | 424 (5.1) | 10 (2.4) |
| General and thoracic epidural only | 10 (0.1) | 0 (0) |
| General and nerve block only | 424 (5.1) | 2 (0.5) |
| Other | 903 (10.8) | 12 (1.3) |

Among 8385 patients, 34.6% (n=2898) underwent a major hip or pelvic surgery, 34.3% (n=2876) underwent knee arthroplasty, 16.8% (n=1405) underwent a major spine surgery, 8.9% (n=750) underwent internal fixation of femur, 3% (n=252) underwent lower leg amputation, and 2.6% (n=221) underwent above-knee amputation. The majority of patients (n=7370, 87.9%) underwent nonemergent surgery.

### Incidence of Death Within 30 Days After Surgery

There were 132 (1.6%) deaths in the orthopedic surgery cohort within 30 days after surgery (Table 2). Of the 132 deaths, 63.6% (84/132) occurred in hospital during the index hospitalization, while the remaining 36.4% (48/132) occurred after discharge. In 1.5% (2/132) of the deaths, patients died in the operating room. The median time to death was 13.5 (IQR 6 - 21) days. Refer to Table 2 for incidence of death according to subcategories of orthopedic surgery. Death rates by nonemergent, urgent, and emergent surgeries were 1.1% (81/7288), 5.7% (46/812), and 3.3% (5/153), respectively. Risk of 30-day mortality was increased in emergent and urgent surgeries compared to nonemergent surgeries in adjusted analyses (HR 2.30, 95% CI 1.61 - 3.42; *P*<.001).

**Table .** Thirty-day perioperative complications, overall and by subtypes of orthopedic surgery[a].

| Outcome | Patients, n (%) | | | | | | |
|---|---|---|---|---|---|---|---|
| | All orthopedic surgery (N=8385) | Subtypes of orthopedic surgery | | | | | |
| | | Major hip or pelvic surgery (n=2898) | Internal fixation of femur (n=750) | Knee arthroplasty (n=2876) | Above-knee amputation (n=221) | Lower leg amputation (n=252) | Major spine surgery (n=1405) |
| Major bleeding | 2422 (28.9) | 965 (33.3) | 287 (38.3) | 725 (25.2) | 103 (46.6) | 96 (38.1) | 258 (18.4) |
| MINS[b] | 1454 (17.3) | 563 (19.4) | 165 (22) | 347 (12.1) | 95 (43) | 95 (37.7) | 197 (14) |
| Sepsis | 318 (3.8) | 124 (4.3) | 33 (4.4) | 53 (1.8) | 30 (13.6) | 20 (7.9) | 60 (4.3) |
| Infection without sepsis | 562 (6.7) | 251 (8.7) | 63 (8.4) | 153 (5.3) | 14 (6.3) | 29 (11.5) | 54 (3.8) |
| Acute kidney injury with dialysis | 17 (0.2) | 4 (0.1) | 3 (0.4) | 2 (0.1) | 2 (0.9) | 3 (1.2) | 3 (0.2) |
| Stroke | 27 (0.3) | 9 (0.3) | 6 (0.8) | 6 (0.2) | 1 (0.5) | 2 (0.8) | 3 (0.2) |
| Venous thromboembolism | 123 (1.5) | 29 (1) | 16 (2.1) | 69 (2.4) | 1 (0.5) | 0 (0) | 8 (0.6) |
| Congestive heart failure | 124 (1.5) | 50 (1.7) | 29 (3.9) | 23 (0.8) | 13 (5.9) | 6 (2.4) | 4 (0.3) |
| New, clinically important atrial fibrillation | 93 (1.1) | 46 (1.6) | 13 (1.7) | 28 (1) | 2 (0.9) | 1 (0.4) | 4 (0.3) |
| Death | 132 (1.6) | 49 (1.7) | 29 (3.9) | 7 (0.2) | 30 (13.6) | 9 (3.6) | 8 (0.6) |

[a]Subtypes of orthopedic surgery were not mutually exclusive (ie, 1 patient could have been categorized to have undergone more than 1 subtype of orthopedic surgery).

[b]MINS: myocardial injury after noncardiac surgery.

### Postoperative Complications and Relationship with 30-Day Mortality

The most common complication was major bleeding (2422/8385, 28.9% patients), followed by MINS (1454/8385, 17.3% patients), infection without sepsis (562/8385, 6.7% patients), sepsis (318/8385, 3.8% patients), then VTE (123/8385, 1.5% patients), congestive heart failure (124/8385, 1.5% patients), new atrial fibrillation (93/8385, 1.1% patients), stroke (27/8385, 0.3% patients), and AKI resulting in dialysis (17/8385, 0.2% patients). Refer to Table 2 for details. The median time from surgery to major bleeding was 2 (IQR 1 - 3) days, MINS 2 (IQR 1 - 3) days, infection without sepsis 10 (IQR 5 - 17) days, and sepsis 8 (IQR 4 - 13) days. Tables 3 and 4 demonstrate the proportion of patients that underwent surgery for an acute fracture as well as the mortality rates according to the timing of surgery performed within each subcategory of orthopedic surgery.

**Table .** Surgery for acute fracture according to subcategories of orthopedic surgery[a].

| Subcategory of orthopedic surgery | Surgery for acute fracture[b], n (%) | |
|---|---|---|
| | No (n=6767) | Yes (n=1616) |
| Major hip or pelvic surgery[c] | 1982 (29.3) | 916 (56.7) |
| Internal femoral fixation | 119 (1.8) | 622 (38.5) |
| Knee arthroplasty | 2854 (42.2) | 21 (1.3) |
| Above-knee amputation | 216 (3.2) | 3 (0.2) |
| Lower leg amputation | 243 (3.6) | 6 (0.4) |
| Major spine surgery | 1353 (20) | 48 (3) |

[a]Surgery for acute fracture was considered nonelective surgery.

[b]2 patients with missing subcategory of orthopedic surgery.

[c]Surgery for hip and pelvis were not mutually exclusive.

**Table .** Mortality rates according to timing of surgery (emergent or urgent vs nonemergent surgery) within each subcategory of orthopedic surgery.

| Subcategory and timing of surgery | Proportions[a], n (%)[b] | Mortality, n | Mortality rate[c], % (95% CI) |
|---|---|---|---|
| Major hip or pelvic surgery (N=2898) | | | |
| Emergent or urgent | 525 (18.1) | 22 | 4.2 (2.7 - 6.4) |
| Nonemergent | 2373 (81.9) | 27 | 1.13 (0.8 - 1.7) |
| Internal femoral fixation (N=741) | | | |
| Emergent or urgent | 313 (42.2) | 14 | 4.5 (2.6 - 7.6) |
| Nonemergent | 428 (57.8) | 15 | 3.5 (2.1 - 5.8) |
| Knee arthroplasty (N=2875) | | | |
| Emergent or urgent | 35 (1.2) | 0 | 0 |
| Nonemergent | 2840 (98.8) | 7 | 0.24 (0.1 - 0.5) |
| Above-knee amputation (N=219) | | | |
| Emergent or urgent | 48 (21.9) | 11 | 23 (12.5 - 37.7) |
| Nonemergent | 171 (78.1) | 19 | 11.2 (7.0 - 17.0) |
| Lower leg amputation (N=249) | | | |
| Emergent or urgent | 49 (19.7) | 3 | 6 (1.6 - 17.9) |
| Nonemergent | 200 (80.3) | 6 | 3.0 (1.2 - 6.7) |
| Major spine surgery (N=1403) | | | |
| Emergent or urgent | 46 (3.3) | 1 | 2 (0.1 - 13.0) |
| Nonemergent | 1357 (96.7) | 7 | 0.51 (0.2 - 1.1) |
| All (N=8385) | | | |
| Emergent or urgent | 1016 (12.1) | 51 | 5.01 (3.8 - 6.6) |
| Nonemergent | 7369 (87.9) | 81 | 1.09 (0.9 - 1.4) |

[a]2 patients with missing subcategory of orthopedic surgery

[b]The proportion was calculated using the total number of patients within each surgical subcategory as the denominator.

[c]The proportion is out of total patients receiving urgent or emergent surgery versus nonemergent, within each subcategory of surgery.

Postoperative complications associated with 30-day mortality included major bleeding (adjusted hazard ratio [aHR] 1.95, 95% CI 1.34 - 2.85), MINS (aHR 2.08, 95% CI 1.38 - 3.14), sepsis (aHR 6.24, 95% CI 3.85 - 10.12), infection without sepsis (aHR 2.74, 95% CI 1.54 - 4.85), stroke (aHR 6.01, 95% CI 2.19 - 16.56), and new clinically important atrial fibrillation (aHR 2.65, 95% CI 1.25 - 5.65). Refer to Table 5 for details. The c-statistic for model performance before and after correction for optimism was 0.87 and 0.85, respectively. Figure 2 is a cumulative hazard curve of the postoperative complications associated with 30-day mortality.

**Table .** Relation between perioperative complications and 30-day mortality in orthopedic surgery.

| Outcome | Patients who died, % (n) (95% CI) | Adjusted HR[a,b] (95% CI) | Attributable fraction (%) |
|---|---|---|---|
| Major bleeding (n=2422) | 3.1 (75) (2.44 - 3.87) | 1.95 (1.34 - 2.85) | 16.5 |
| No major bleeding (n=5960) | 0.9 (57) (0.73 - 1.24) | Reference | N/A[c] |
| MINS[d] (n=1454) | 4.3 (63) (3.35 - 5.51) | 2.08 (1.38 - 3.14) | 20.6 |
| No MINS (n=6931) | 0.9 (69) (0.78 - 1.26) | Reference | N/A |
| Sepsis (n=318) | 9.4 (30) (6.46 - 13.19) | 6.24 (3.85 - 10.12) | 9.7 |
| Infection without sepsis (n=562) | 3 (17) (1.77 - 4.80) | 2.74 (1.54 - 4.85) | 3.8 |
| No sepsis or infection (n=7503) | 1.1 (85) (0.91 - 1.40) | Reference | N/A |
| Acute kidney injury with dialysis (n=17) | 41.2 (7) (18.44 - 67.07) | N/A | N/A |
| No acute kidney injury with dialysis (n=8366) | 1.5 (125) (1.25 - 1.78) | N/A | N/A |
| Stroke (n=27) | 14.8 (4) (4.19 - 33.73) | 6.01 (2.19 - 16.56) | 1.5 |
| No stroke (n= 8356) | 1.5 (128) (1.28 - 1.82) | Reference | N/A |
| Venous thromboembolism (n=123) | 2.4 (3) (0.51 - 6.96) | 2.24 (0.70 - 7.13) | N/A |
| No venous thromboembolism (n=8261) | 1.6 (129) (1.31 - 1.85) | Reference | N/A |
| Congestive heart failure (n=124) | 15.3 (19) (9.48 - 22.89) | 1.54 (0.81 - 2.94) | N/A |
| No congestive heart failure (n=8259) | 1.4 (13) (1.13 - 1.64) | Reference | N/A |
| New, clinically important atrial fibrillation (n=93) | 10.8 (10) (5.28 - 18.89) | 2.65 (1.25 - 5.65) | 2.2 |
| No new, clinically important atrial fibrillation (n=8290) | 1.5 (122) (1.22 - 1.75) | Reference | N/A |

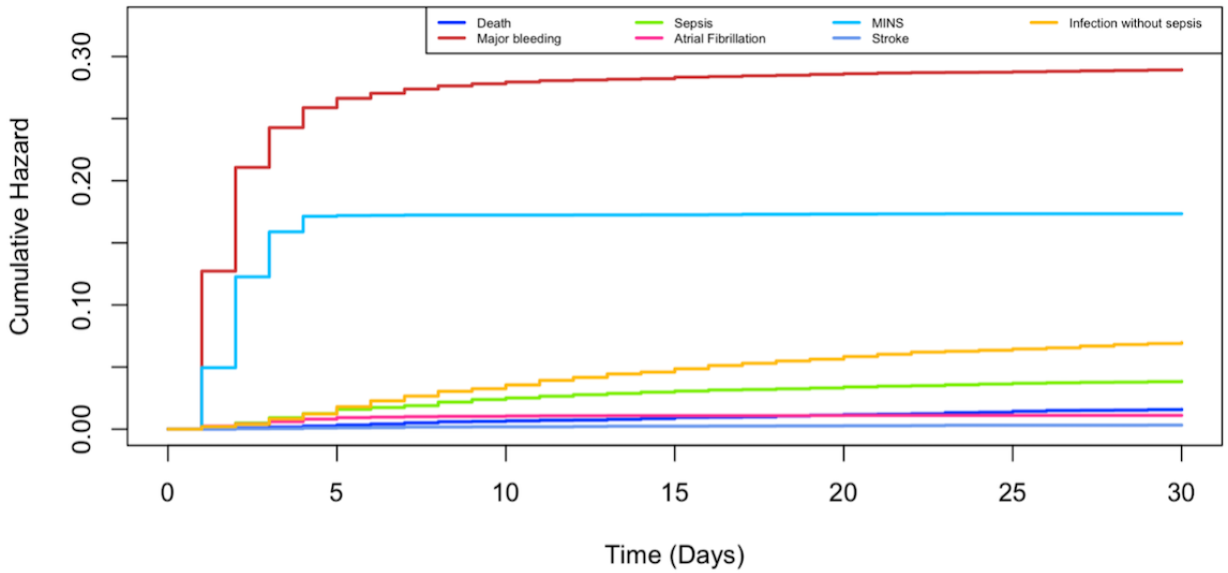
[a]HR: hazard ratio.

[b]Adjusted variables were as follows: age, history of peripheral vascular disease, history of chronic obstructive pulmonary disease, surgery urgency, and active cancer.

[c]N/A: not applicable.

[d]MINS: myocardial injury after noncardiac surgery.

**Figure 2.** Cumulative hazard curve of postoperative complications associated with 30-day mortality in orthopedic surgery. MINS: myocardial injury after noncardiac surgery.




Among the postoperative complications significantly associated with mortality, the greatest attributable fraction was for MINS (20.6%), followed by major bleeding (16.5%), then sepsis (9.7%), infection without sepsis (3.8%), new atrial fibrillation (2.2%), and stroke (1.5%).

Post hoc subgroup analyses performed according to surgery urgency demonstrated similar findings. Among those who underwent nonemergent surgeries, there were 81 deaths. Following adjustment, MINS (HR 2.02, 95% CI 1.22 - 3.32), atrial fibrillation (HR 2.64, 95% CI 1.04 - 6.89), major bleeding (HR 2.30, 95% CI 1.41 - 3.76), sepsis (HR 6.27, 95% CI 3.43 - 11.46), and stroke (HR 10.49, 95% CI 3.18 - 34.63) were significantly associated with 30-day mortality. Among those who underwent urgent and emergent surgeries, there were 51 deaths. In this subgroup, with further reduced number of mortality events, infection without sepsis (HR 3.49, 95% CI 1.85 - 6.24), MINS (HR 3.40, 95% CI 1.85 - 6.24), and sepsis (HR 5.65, 95% CI 2.51 - 12.70) were significantly associated with 30-day mortality.

## Discussion

### Principal Findings

This large international prospective cohort study involving 8385 patients who underwent orthopedic surgery demonstrated a 1.6% (132/8385) mortality rate. Major bleeding, MINS, infection without sepsis, sepsis, stroke, and new clinically important atrial fibrillation were significantly associated with 30-day mortality. Among these, the greatest attributable fraction for death in our cohort was from MINS (1454/8385, 21%), major bleeding (2422/8385, 17%), sepsis (318/8385, 10%), infection without sepsis (562/8385, 4%), stroke (27/8385, 2%), and new atrial fibrillation (93/8385, 2%).

Consistent with prior studies that reported 30-day mortality rates ranging from 0.6% to 8.7%, this work demonstrated a 1.6% (132/8385) mortality rate [2-4]. Our study also demonstrated substantial variations in mortality across subcategories of orthopedic surgery and surgery urgency. Traditional preoperative risk scoring systems do not consider the nuances between surgical procedures that contribute to notable variations in mortality risk [25]. We found substantial mortality differences across surgical procedures within orthopedic surgery, which urges greater consideration of procedure type for preoperative risk stratification and consideration of higher-level postoperative monitoring (eg, admission to intensive cardiac care unit and telemetry).

Among the few studies that investigate complications associated with mortality in orthopedic surgery, pneumonia, acute renal failure, stroke, myocardial infarction, and MINS have been reported [5,26]. Our results are unique in identifying major bleeding to be a risk factor for death in the orthopedic surgery cohort. Previous studies did not select postoperative bleeding as a potential variable to explore, despite 80% of 739 orthopedic surgeons and 50 anesthesiologists in an international survey responding that they were either concerned or very concerned for bleeding in orthopedic surgery populations [27]. Our results demonstrate that major bleeding in orthopedic surgery is common and is associated with mortality, necessitating further investigations in this area to improve patient outcomes.

Although a major bleeding event was the most common complication in our cohort, MINS demonstrated the greatest attributable fraction for death. In other words, although an overall smaller number of patients experienced MINS, a greater number of patients with MINS died within 30 days of surgery compared to the number of patients who experienced major bleeding. This is unique to the orthopedic surgery cohort compared to the overall VISION mortality analyses and general surgery mortality substudy, where major bleeding had the highest attributable fraction for death [11,17]. Certainly, the varying risk profiles, propensity for bleeding, and cardiovascular stress by surgical subspecialties contribute to the differences in

these results. This emphasizes the need for specialty-specific data to inform surgery-specific practice.

Another potential explanation is the unique practice of bleeding prophylaxis that is commonplace in orthopedic surgery [28-31]. Consistent demonstration of perioperative bleeding reduction with tranexamic acid use in orthopedic surgery–specific research has led to the routine use of prophylactic tranexamic acid in orthopedic surgery for many years, particularly in trauma, joint, and spine surgery [28-33]. This is not commonplace in other noncardiac surgical specialties, despite recent evidence to support its safety and efficacy in these contexts [28-30,34-37]. The long-standing practice of routine prophylaxis for perioperative bleeding in orthopedic surgery may have contributed to reduced bleeding severity and lower attributable fraction for death in this cohort compared to the overall and general surgery cohorts. If this were to be true, it is also worth noting that the rate of VTE in this orthopedic cohort was only around 1% (123/8385), which is similar to the incidence found in the noncardiac and general surgery cohorts [11].

Of the postoperative complications investigated, MINS was the second most common complication occurring in 17.3% (1454/8385) of the orthopedic surgery cohort. We also demonstrated a 12% (347/2876) incidence of MINS in the knee arthroplasty subcategory. Previously reported rates of myocardial infarction (MI) after knee arthroplasty have ranged between 0.3% and 2.2% [38,39]. Considering MINS encompasses a broader spectrum of myocardial injury than MI, the larger incidence of MINS was expected, especially as high-sensitivity troponin assays were used for monitoring. Among the few existing studies investigating the incidence of MINS, the results have varied significantly, likely due to limitations in very small sample sizes (ie, 1/82, 1.2% incidence vs 68/160, 42% incidence) [40,41]. Our study included 2876 patients who underwent knee arthroplasty, representing the largest cohort in which the epidemiology of MINS has been studied. As MINS is largely asymptomatic, these findings highlight the importance of identifying MINS for postoperative risk stratification and downstream management, as emerging evidence suggests targeted treatment may reduce future cardiovascular events [12,16,25,42].

Finally, subgroup analyses by surgery urgency demonstrated minor differences in postoperative complications that were found to be associated with mortality between the emergent and nonemergent groups. Specifically, major bleeding was not associated with mortality in the urgent-emergent subgroup. Clinical interpretation of these findings is limited as these are post hoc analyses with reduced events per variable and therefore, model instability. However, these findings suggest differences in mortality determinants highlighting tailored strategies to address perioperative risks according to surgery urgency.

## Limitations

There are a few limitations to consider. Only patients aged ≥45 years were included; thus, these findings may not apply to younger patients. Baseline variables (ie, age and estimated glomerular filtration rate) were modeled as a categorical variable to align with prior VISION analyses, although this approach may not fully capture potential nonlinear relationships. Additionally, as our inclusion criteria were targeted to include a higher risk surgical population, the present results would not apply for patients undergoing low-risk, same-day surgeries. The granularity of the data are limited to the subcategories of orthopedic surgery that were predefined before the start of the study [43-45]. Although orthopedic procedures are heterogeneous, combining them enabled sufficient power to evaluate associations between relatively infrequent complications and mortality. We report subtype-specific complication rates (Table 2), demonstrating that major bleeding, MINS, and sepsis were consistently among the most common complications across procedures. While procedure-specific attributable fractions were not feasible due to limited events, our findings identify broadly relevant targets for perioperative risk reduction. The stability of a Cox regression model is dependent on the events per variable. The base model meets the lower acceptable cutoff at 10 events per variable, suggesting adequate power for the primary analysis; there remains potential risk for overfitting [24,46]. However, the negligible change in the c-statistic after correcting for optimism suggests overfitting may not be a significant issue [20]. Variation in VTE prophylaxis practices, which were not captured in detail in this study, may contribute to differences in bleeding risk and represents a potential unmeasured confounder. Finally, complications such as infection and sepsis may occur later in the postoperative course, potentially reducing their relative contribution in time-dependent analyses within a 30-day follow-up period, particularly given their lower event frequency compared to earlier complications such as bleeding or MINS.

To our knowledge, this is the first large prospective study exploring the association of various postoperative complications and death in a global prospective orthopedic surgery study. The inclusion of diverse participants across 14 countries increases the generalizability of our results. Furthermore, with no loss to follow-up and the collection of time-dependent data, we were able to increase precision and reduce bias in fulfilling our objectives by creating a realistic model that accounts for the changes in risk factors for survival that occur over time.

## Conclusions

This large international prospective cohort study of patients undergoing orthopedic surgery (7369/8385, 88% nonemergent) demonstrated a 30-day mortality rate of 1.6% (132/8385). Adjusted analyses demonstrated major bleeding, MINS, sepsis, infection without sepsis, stroke, and atrial fibrillation to be associated with mortality. The highest attributable fraction of death in our cohort was contributed by MINS, major bleeding, sepsis, and infection without sepsis, which highlights areas for further study to reduce mortality among patients who underwent orthopedic surgery.

## Funding

Please see Multimedia Appendix 1 for funding sources.

## Data Availability

The datasets generated or analyzed during this study are not publicly available due but are available from the corresponding author on reasonable request.

## Conflicts of Interest

On the basis of study questions PJD has originated and grants he has written, he has received grants from Abbott Diagnostics, AOP Pharma, AstraZeneca, Bayer, Boehringer Ingelheim, Bristol-Myers-Squibb, CloudDX, Covidien, Octapharma, Philips Healthcare, Roche Diagnostics, Siemens, and Stryker. PJD has participated in advisory board meetings for GlaxoSmithKline, Boehringer Ingelheim, Bayer, and Quidel Canada. He attended expert panel meetings with AstraZeneca and Boehringer Ingelheim; served as a consultant for a call with Roche Pharma; and conducted consultancy work for Abbott, AstraZeneca, Renibus Therapeutics, Roche Canada, and Trimedic. He has also been invited as a speaker with Bayer Inc, Novartis Pharma Canada, and Abbott Diagnostics. FKB holds a PHRI Career Award and a McMaster University Department of HEI Mid-Career Research Award. She received investigator-initiated research grants from Roche Diagnostics and SIEMENS, and attended advisory board meetings from Abbott. ED received investigator-initiated research grants from Roche Laboratories and Abbott Laboratories, outside of the submitted work. She receives a clinician-researcher salary award from the Fonds de recherche du Québec-Santé. DC received consulting fees from Roche Diagnostics and Trimedics. PES has received consulting fees from Incyte Biosciences Canada and Hoffman La Roche. ORA reports a relationship with Stryker Canada that includes speaking and lecture fees. He is the tier 2 Canada Research Chair in Joint Preservation and President/Owner of Notch Academy. MKW is supported by the CANTRAIN Clinical Trials Training Platforms – Doctoral Studentship Award. DT has received funding from Stryker Canada.

Multimedia Appendix 1
VISION funding sources.
[DOCX File, 17 KB - periop_v9i1e90823_app1.docx ]

Multimedia Appendix 2
VISION postoperative complications and baseline variable definitions.
[DOCX File, 20 KB - periop_v9i1e90823_app2.docx ]

Checklist 1
STROBE checklist.
[DOCX File, 22 KB - periop_v9i1e90823_app3.docx ]

## References

1. Kamel I, Ahmed MF, Sethi A. Regional anesthesia for orthopedic procedures: what orthopedic surgeons need to know. World J Orthop 2022 Jan 18;13(1):11-35. [doi: 10.5312/wjo.v13.i1.11] [Medline: 35096534]
2. Banerjee S, Suresh G, Kale AB, Sathe AH. An audit of admissions and mortality of orthopedic indoor patients in a tertiary care hospital of India. J Clin Orthop Trauma 2020 Jul;11(Suppl 4):S518-S521. [doi: 10.1016/j.jcot.2020.04.012] [Medline: 32774021]
3. Sheikh HQ, Hossain FS, Aqil A, Akinbamijo B, Mushtaq V, Kapoor H. A comprehensive analysis of the causes and predictors of 30-day mortality following hip fracture surgery. Clin Orthop Surg 2017 Mar;9(1):10-18. [doi: 10.4055/cios.2017.9.1.10] [Medline: 28261422]
4. Tan HB, MacDonald DA, Matthews SJ, Giannoudis PV. Incidence and causes of mortality following acute orthopaedic and trauma admissions. Ann R Coll Surg Engl 2004 May;86(3):156-160. [doi: 10.1308/003588404323043256] [Medline: 15140297]
5. Bhattacharyya T, Iorio R, Healy WL. Rate of and risk factors for acute inpatient mortality after orthopaedic surgery. J Bone Joint Surg Am 2002 Apr;84(4):562-572. [doi: 10.2106/00004623-200204000-00009] [Medline: 11940616]
6. Ramirez LF, Thisted R. Using a national health care data base to determine surgical complications in community hospitals: lumbar discectomy as an example. Neurosurgery 1989 Aug;25(2):218-225. [doi: 10.1097/00006123-198908000-00011] [Medline: 2770986]
7. Ciol MA, Deyo RA, Howell E, Kreif S. An assessment of surgery for spinal stenosis: time trends, geographic variations, complications, and reoperations. J Am Geriatr Soc 1996 Mar;44(3):285-290. [doi: 10.1111/j.1532-5415.1996.tb00915.x] [Medline: 8600197]
8. Derkay CS. Pediatric otolaryngology procedures in the United States: 1977-1987. Int J Pediatr Otorhinolaryngol 1993 Jan;25(1-3):1-12. [doi: 10.1016/0165-5876(93)90004-m] [Medline: 8436452]

9. Gillum RF. Peripheral arterial occlusive disease of the extremities in the United States: hospitalization and mortality. Am Heart J 1990 Dec;120(6 Pt 1):1414-1418. [doi: 10.1016/0002-8703(90)90257-x] [Medline: 2248187]
10. Owings MF, Kozak LJ. Ambulatory and inpatient procedures in the United States, 1996. Vital Health Stat 13 1998 Nov;13(139):1-119 [FREE Full text] [Medline: 9866429]
11. Vascular Events in Noncardiac Surgery Patients Cohort Evaluation (VISION) Study Investigators, Spence J, LeManach Y, et al. Association between complications and death within 30 days after noncardiac surgery. CMAJ 2019 Jul 29;191(30):E830-E837. [doi: 10.1503/cmaj.190221] [Medline: 31358597]
12. Thomas S, Borges F, Bhandari M, et al. Association between myocardial injury and cardiovascular outcomes of orthopaedic surgery: a vascular events in noncardiac surgery patients cohort evaluation (VISION) substudy. J Bone Joint Surg Am 2020 May 20;102(10):880-888. [doi: 10.2106/JBJS.18.01305] [Medline: 32118652]
13. Salerno SM, Carlson DW, Soh EK, Lettieri CJ. Impact of perioperative cardiac assessment guidelines on management of orthopedic surgery patients. Am J Med 2007 Feb;120(2):185. [doi: 10.1016/j.amjmed.2005.11.009] [Medline: 17275461]
14. Devereaux PJ, Goldman L, Cook DJ, Gilbert K, Leslie K, Guyatt GH. Perioperative cardiac events in patients undergoing noncardiac surgery: a review of the magnitude of the problem, the pathophysiology of the events and methods to estimate and communicate risk. CMAJ 2005 Sep 13;173(6):627-634. [doi: 10.1503/cmaj.050011] [Medline: 16157727]
15. Vascular Events In Noncardiac Surgery Patients Cohort Evaluation (VISION) Study Investigators, Devereaux PJ, Chan MT, et al. Association between postoperative troponin levels and 30-day mortality among patients undergoing noncardiac surgery. JAMA 2012 Jun 6;307(21):2295-2304. [doi: 10.1001/jama.2012.5502] [Medline: 22706835]
16. Writing Committee for the VISION Study Investigators, Devereaux PJ, Biccard BM, 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]
17. Park LJ, Borges FK, Ofori S, et al. Association between complications and death within 30 days after general surgery: a vascular event in noncardiac surgery patients cohort evaluation (VISION) substudy. Ann Surg 2025 Dec 1;282(6):1007-1013. [doi: 10.1097/SLA.0000000000006372] [Medline: 38860367]
18. Roshanov PS, Eikelboom JW, Crowther M, et al. Bleeding impacting mortality after noncardiac surgery: a protocol to establish diagnostic criteria, estimate prognostic importance, and develop and validate a prediction guide in an international prospective cohort study. CMAJ Open 2017 Aug 4;5(3):E594-E603. [doi: 10.9778/cmajo.20160106] [Medline: 28943515]
19. Roshanov PS, Eikelboom JW, Sessler DI, et al. Bleeding Independently associated with Mortality after noncardiac Surgery (BIMS): an international prospective cohort study establishing diagnostic criteria and prognostic importance. Br J Anaesth 2021 Jan;126(1):163-171. [doi: 10.1016/j.bja.2020.06.051] [Medline: 32768179]
20. Smith GC, Seaman SR, Wood AM, Royston P, White IR. Correcting for optimistic prediction in small data sets. Am J Epidemiol 2014 Aug 1;180(3):318-324. [doi: 10.1093/aje/kwu140] [Medline: 24966219]
21. Rückinger S, von Kries R, Toschke AM. An illustration of and programs estimating attributable fractions in large scale surveys considering multiple risk factors. BMC Med Res Methodol 2009 Jan 23;9(1):7. [doi: 10.1186/1471-2288-9-7] [Medline: 19166593]
22. von Cube M, Timsit JF, Schumacher M, Motschall E, Schumacher M. Quantification and interpretation of attributable mortality in core clinical infectious disease journals. Lancet Infect Dis 2020 Dec;20(12):e299-e306. [doi: 10.1016/S1473-3099(20)30485-0] [Medline: 32916101]
23. Mansournia MA, Altman DG. Population attributable fraction. BMJ 2018 Feb 22;360:k757. [doi: 10.1136/bmj.k757] [Medline: 29472187]
24. Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and Cox regression. Am J Epidemiol 2007 Mar 15;165(6):710-718. [doi: 10.1093/aje/kwk052] [Medline: 17182981]
25. Duceppe E, Parlow J, MacDonald P, et al. Canadian Cardiovascular Society guidelines on perioperative cardiac risk assessment and management for patients who undergo noncardiac surgery. Can J Cardiol 2017 Jan;33(1):17-32. [doi: 10.1016/j.cjca.2016.09.008] [Medline: 27865641]
26. Hu W, Chen Y, Zhao K, et al. Association of perioperative myocardial injury with 30-day and long-term mortality in older adult patients undergoing orthopedic surgery in China. Med Sci Monit 2021 Dec 25;27:e932036. [doi: 10.12659/MSM.932036] [Medline: 34952895]
27. Ginzburg E, Dujardin F. Physicians' perceptions of the definition of major bleeding in major orthopedic surgery: results of an international survey. J Thromb Thrombolysis 2011 Feb;31(2):188-195. [doi: 10.1007/s11239-010-0498-9] [Medline: 20640486]
28. Wong J, George RB, Hanley CM, Saliba C, Yee DA, Jerath A. Tranexamic acid: current use in obstetrics, major orthopedic, and trauma surgery. Can J Anesth 2021 Jun;68(6):894-917. [doi: 10.1007/s12630-021-01967-7] [Medline: 33993459]
29. Jennings JD, Solarz MK, Haydel C. Application of tranexamic acid in trauma and orthopedic surgery. Orthop Clin North Am 2016 Jan;47(1):137-143. [doi: 10.1016/j.ocl.2015.08.014] [Medline: 26614928]
30. Lin ZX, Woolf SK. Safety, efficacy, and cost-effectiveness of tranexamic acid in orthopedic surgery. Orthopedics 2016;39(2):119-130. [doi: 10.3928/01477447-20160301-05] [Medline: 26942474]
31. Colomina MJ, Contreras L, Guilabert P, Koo M, M Ndez E, Sabate A. Clinical use of tranexamic acid: evidences and controversies. Braz J Anesthesiol 2022;72(6):795-812. [doi: 10.1016/j.bjane.2021.08.022] [Medline: 34626756]

32. Gandhi R, Evans HM, Mahomed SR, Mahomed NN. Tranexamic acid and the reduction of blood loss in total knee and hip arthroplasty: a meta-analysis. BMC Res Notes 2013 May 7;6:184. [doi: 10.1186/1756-0500-6-184] [Medline: 23651507]
33. Goldstein K, Jones C, Kay J, Shin J, de Sa D. Tranexamic acid administration in arthroscopic surgery is a safe adjunct to decrease postoperative pain and swelling: a systematic review and meta-analysis. Arthroscopy 2022 Apr;38(4):1366-1377. [doi: 10.1016/j.arthro.2021.10.001] [Medline: 34655767]
34. Montroy J, Hutton B, Fergusson DA, et al. Lysine analogue use during cancer surgery: a survey from a Canadian tertiary care centre. Curr Oncol 2020 Dec;27(6):e560-e568. [doi: 10.3747/co.27.6613] [Medline: 33380871]
35. Devereaux PJ, Marcucci M, Painter TW, et al. Tranexamic acid in patients undergoing noncardiac surgery. N Engl J Med 2022 May 26;386(21):1986-1997. [doi: 10.1056/NEJMoa2201171] [Medline: 35363452]
36. Tsan SE, Viknaswaran NL, Cheong CC, et al. Prophylactic intravenous tranexamic acid and thromboembolism in non-cardiac surgery: a systematic review, meta-analysis and trial sequential analysis. Anaesthesia 2023 Sep;78(9):1153-1161. [doi: 10.1111/anae.16058] [Medline: 37314744]
37. Houston BL, Fergusson DA, Falk J, et al. Prophylactic tranexamic acid use in non-cardiac surgeries at high risk for transfusion. Transfus Med 2021 Aug;31(4):236-242. [doi: 10.1111/tme.12780] [Medline: 33938051]
38. Lu N, Misra D, Neogi T, Choi HK, Zhang Y. Total joint arthroplasty and the risk of myocardial infarction: a general population, propensity score–matched cohort study. Arthritis Rheumatol 2015 Oct;67(10):2771-2779. [doi: 10.1002/art.39246] [Medline: 26331443]
39. Petersen PB, Kehlet H, Jørgensen CC, Lundbeck Foundation Center for Fast-track Hip and Knee Replacement Collaborative Group. Myocardial infarction following fast-track total hip and knee arthroplasty-incidence, time course, and risk factors: a prospective cohort study of 24,862 procedures. Acta Orthop 2018 Dec;89(6):603-609. [doi: 10.1080/17453674.2018.1517487] [Medline: 30326755]
40. Van Zyl RD, Burger MC, Jordaan JD. Prevalence of a postoperative troponin leak in patients with cardiac risk factors undergoing knee and hip arthroplasty in a South African population. S Afr Med J 2020 Mar 30;110(4):320-326. [doi: 10.7196/SAMJ.2020.v110i4.14133] [Medline: 32657745]
41. Lee S, Allen AJ, Morley E, et al. Perioperative myocardial injury risk after elective knee and hip arthroplasty in patients with a high risk of obstructive sleep apnea. Sleep Breath 2021 Mar;25(1):513-515. [doi: 10.1007/s11325-020-02068-3] [Medline: 32253610]
42. Devereaux PJ, Duceppe E, Guyatt G, et al. Dabigatran in patients with myocardial injury after non-cardiac surgery (MANAGE): an international, randomised, placebo-controlled trial. Lancet 2018 Jun 9;391(10137):2325-2334. [doi: 10.1016/S0140-6736(18)30832-8] [Medline: 29900874]
43. Thakker A, Briggs N, Maeda A, Byrne J, Davey JR, Jackson TD. Reducing the rate of post-surgical urinary tract infections in orthopedic patients. BMJ Open Qual 2018;7(2):e000177. [doi: 10.1136/bmjoq-2017-000177] [Medline: 29719874]
44. Al-Mulhim FA, Baragbah MA, Sadat-Ali M, Alomran AS, Azam MQ. Prevalence of surgical site infection in orthopedic surgery: a 5-year analysis. Int Surg 2014;99(3):264-268. [doi: 10.9738/INTSURG-D-13-00251.1] [Medline: 24833150]
45. Seidelman JL, Mantyh CR, Anderson DJ. Surgical site infection prevention: a review. JAMA 2023 Jan 17;329(3):244-252. [doi: 10.1001/jama.2022.24075] [Medline: 36648463]
46. Ogundimu EO, Altman DG, Collins GS. Adequate sample size for developing prediction models is not simply related to events per variable. J Clin Epidemiol 2016 Aug;76:175-182. [doi: 10.1016/j.jclinepi.2016.02.031] [Medline: 26964707]


## Abbreviations

**aHR:** adjusted hazard ratio
**AKI:** acute kidney injury
**HR:** hazard ratio
**MI:** myocardial infarction
**MINS:** myocardial injury after noncardiac surgery
**VISION:** Vascular Events in Noncardiac Surgery Patients Cohort Evaluation
**VTE:** venous thromboembolism

*Edited by J Shiffermiller; submitted 04.Jan.2026; peer-reviewed by M Pappas, S Cohn; revised version received 28.Apr.2026; accepted 01.May.2026; published 09.Jun.2026.*

*Please cite as:*
*Park LJ, Devereaux PJ, Patel A, Tandon V, Heels-Ansdell D, Thabane L, Serrano PE, Chan MTV, Szczeklik W, Srinathan S, Garutti I, Urrutia G, Guerra-Farfan E, Abdel Khalik H, Duceppe E, Ofori S, Marcucci M, Conen D, Wang MK, Spence J, Tushinski D, Bali K, Adili A, Khanna V, Tonelli AC, Mulazzani F, Jiang W, Ayeni OR, Slobegean G, Miclau T, Bhandari M, Borges FK*
*Association Between Complications and Death Within 30 Days After Orthopedic Surgery: Vascular Events in Noncardiac Surgery Patients Cohort Evaluation (VISION) Substudy*
*JMIR Perioper Med 2026;9:e90823*
*URL: https://periop.jmir.org/2026/1/e90823*
*doi:10.2196/90823*

# Assessing the Effects of eHealth Literacy and the Area Deprivation Index on Barriers to Electronic Patient Portal Use for Orthopedic Surgery: Cross-Sectional Observational Study

Audrey Lynn Litvak[1*], BA; Nicholas Lin[2*], MD; Kelly Hynes[3*], MD; Jason Strelzow[3*], MD; Jeffrey G Stepan[2*], MD, MSc

[1]Pritzker School of Medicine, University of Chicago, 924 E 57th Street #104, Chicago, IL, United States

[2]Department of Orthopaedic Surgery and Rehabilitation Services, University of Chicago, Chicago, IL, United States

[3]Department of Orthopaedic Surgery, Washington University in St Louis, St Louis, MO, United States

[*]all authors contributed equally

**Corresponding Author:**
Audrey Lynn Litvak, BA
Pritzker School of Medicine, University of Chicago, 924 E 57th Street #104, Chicago, IL, United States

## Abstract

**Background:** As electronic patient portals (EPPs) continue to gain popularity and systems transition to online tools for scheduling, communication, and telehealth, patients without access or skills to use these tools may be overlooked.

**Objective:** This study analyzed patient and neighborhood-level factors, including eHealth literacy level and the Area Deprivation Index (ADI), that may limit EPP access for orthopedic surgery.

**Methods:** A cross-sectional, survey-based study was performed at a single urban tertiary academic medical center in the United States across foot and ankle, hand and upper extremity, and orthopedic trauma subspecialty clinics from June 21, 2022, to August 12, 2022. Survey responses (N=287) provided information on sociodemographic characteristics; barriers to EPP use and frequency of EPP use; the eHealth Literacy Scale; and the ADI, which is an address-generated national census measure of neighborhood-level disadvantage. Barriers to EPP use were inductively coded into barrier types, classified as physical access, technology discomfort, or preference. The primary outcome measure was patient-reported barriers to EPP use, which was treated as a binary outcome (1=barrier; 0=no barrier). Bivariate analyses and multivariable binary logistic regressions were performed.

**Results:** The percentage of patients who self-reported barriers to EPP access was 43.2% (124/287), which related to physical access (13/124, 10.4%), technology discomfort (55/124, 44.3%), and preference (78/124, 63.0%). In the adjusted regressions, only low eHealth literacy and older age predicted barriers to EPP use (low eHealth literacy, adjusted odds ratio [AOR] 1.32, 95% CI 1.13-1.54; *P*<.001; older age, AOR 1.007, 95% CI 1.003-1.009; *P*<.001), including barriers of technology discomfort (low eHealth literacy, AOR 1.25, 95% CI 1.11-1.40; *P*<.001; older age, AOR 1.004, 95% CI 1.002-1.007; *P*<.001) and preference (low eHealth literacy, AOR 1.33, 95% CI 1.17-1.51; *P*<.001; older age, AOR 1.004, 95% CI 1.00-1.01; *P*<.01). Patients with physical access–related barriers as opposed to technology discomfort or preference barriers had the lowest median eHealth literacy scores (17.0, IQR 12.0-14.0 vs 27.0, IQR 16.0-32.0 vs 27.0, IQR 20.0-32.0, respectively) and roughly a quartile higher median ADI (73.0, IQR 41.0-92.0 vs 53.5, IQR 31.2-76.0 vs 58.0, IQR 38.8-83.8, respectively).

**Conclusions:** Low eHealth literacy was the most significant determinant of overall barriers to EPP use for orthopedic surgery, followed by older age. Neighborhood-level disadvantage as measured through the ADI had no mediating effect on patient-reported barriers to EPP use when adjusting for eHealth literacy level. While patients with physical access barriers had higher ADIs, overall, few patients reported physical access barriers compared to barriers related to technology discomfort or preference. Patient preference for EPP versus non-EPP communications should be documented. Point-of-care screening using the eHealth Literacy Scale may also identify patients who require follow-up outside of the EPP during critical perioperative periods.

***(JMIR Perioper Med 2026;9:e72035)*** doi:10.2196/72035

**KEYWORDS**

electronic health records; eHealth literacy; online systems; health equity; social determinants of health; SDOH; orthopedic surgery

## Introduction

Improving digital health information transparency and transmission via electronic patient portals (EPPs) has been a central focus of health IT policy in the United States during the last decade [1]. EPPs facilitate patients' communication with their treating care teams and direct access to their electronic personal health record and tools to request prescription refills, participate in e-visits, and complete patient-reported outcome questionnaires, among other functions. These interactions empower patients to take an active role in their health care [1,2]. However, few studies have examined patient portal use in the surgical setting or patient factors that may limit EPP use in this context [3-8].

Benefits of EPP enrollment among orthopedic patients include improved patient outcomes [4,7], medication adherence [3], higher patient satisfaction and psychosocial health [3,7], and fewer missed appointments [3]. Patient engagement via the EPP may additionally facilitate more effective screening for commonly avoidable complications that delay patients' return to function, such as soaking of splints or patient-prolonged immobilization due to unanticipated postoperative pain. Moreover, as routine messaging and completion of patient-reported outcomes via the EPP becomes standard, it is likely that patient engagement via the EPP beyond enrollment may become another critical quality metric tied to physician reimbursement.

Despite advantages of and health care provider interest in adoption of EPP tools, prior studies in orthopedic surgery have shown that patient factors, including older age and lower educational level, may limit EPP enrollment [3,4,6,8], which is analogous to observations in the internal medicine setting [9-12]. The 2020 Health Information National Trends Survey (HINTS) found that the most commonly cited reason for patient nonuse within a large US sample was desire to speak directly with a health care provider (ie, physician, nurse practitioner)—a sentiment shared by 69% of patients [11]. Furthermore, roughly 30% of patients expressed discomfort with the technology [11]. Traditional health literacy refers to the capacity to find, understand, and use health information to inform health-related decisions and actions, whereas eHealth literacy specifically refers to the capacity and skills to seek, assess, and make use of health information via electronic media. In the hospital medicine setting, low eHealth literacy in particular is associated with less awareness, use, and perceived usefulness of EPPs [13].

Lack of examination of granular patient-level use data beyond EPP activation status, explicit barriers to EPP use, and associated patient factors such as health literacy are described as significant limitations and directions for future work in orthopedic surgery [3,4,8]. Additionally, no study across any prior setting has assessed the effect of structural or neighborhood-level determinants on barriers to EPP use, nor have they assessed barriers among patients who are actively enrolled in EPPs. Individual-level determinants may refer to patient demographics or skill sets such as health literacy, whereas neighborhood-level determinants refer to unmeasured social factors conferred by the geographic environment in which a patient lives, often described via census variables related to percentage of unemployment, percentage of individuals with a high school education, and food and housing quality, among others.

For digital health uptake in particular, distinguishing among types and levels of determinants is critical to realizing equity-informed intervention and policy [14,15]. For example, digital literacy is an individual-level factor for which a policy-level solution such as improving broadband connectivity or personal device accessibility may be ineffective in the absence of community-responsive interventions to provide individuals with digital skill training [15]. In particular, it is important to analyze whether neighborhood disadvantage may amplify the impact of low eHealth literacy on barriers to patient portal use, which single-level analyses of eHealth literacy cannot capture.

This study aimed to contribute to the existing body of literature on barriers to EPP use in orthopedic surgery by analyzing how both individual-level and neighborhood-level social determinants, including eHealth literacy and the Area Deprivation Index (ADI), may relate to patient-reported barriers to EPP access and use across foot and ankle, hand and upper extremity, and orthopedic trauma surgery.

## Methods

### Study Design and Setting

This was a cross-sectional survey–based study conducted via an anonymized paper survey administered at a single urban tertiary academic medical center in the United States between June 21, 2022, and August 12, 2022. The survey was administered in the clinic following each patient visit and consisted of sociodemographic questions, the eHealth Literacy Scale (eHEALS), and 2 questions regarding EPP access and use detailed below. This study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) design and reporting guidelines for cross-sectional studies.

### Participants

All English-speaking patients aged >18 years presenting for orthopedic surgery evaluation at foot and ankle, trauma, and hand and upper extremity clinics were included and approached consecutively during the aforementioned period. This study was limited to foot and ankle, trauma, and hand and upper extremity surgery clinics where faculty involvement at our institution was feasible. Approached patients were excluded if they had not received a tablet to complete their in-office patient-reported outcome questionnaire per routine standard of care due to external technological or capacity constraints (ie, tablet out of battery or too few working tablets in the clinic on a particular day) unless that patient had already completed the questionnaire via their patient portal. While this practical limitation resulted in potential selection bias, it randomly affected only a small portion of patients (<10) and was remedied to avoid recurrence of the problem for continued recruitment efforts. Patients who could not read or write were included and read aloud the study survey.

### Sample Size Calculation

The HINTS 2020 reported that approximately 40% of adults in the United States accessed their patient portal in the previous year, whereas 59% were nonusers [11]. Using a baseline nonuse or potential barrier rate of roughly 40% to 60% and setting an α value of .05, we estimated that a sample of 270 to 290 patients would provide at least 80% power to detect large effect sizes (odds ratio 2.0 - 3.0) in a binary regression model constrained by 10 events per covariate included.

### Ethical Considerations

Biological Sciences Division/University of Chicago Institutional Review Board approval (IRB22-0230) was obtained with waivers for written consent and HIPAA (Health Insurance Portability and Accountability Act) authorization. Anonymized survey data were transcribed into a REDCap (Research Electronic Data Capture; Vanderbilt University) database for secure storage [16]. No compensation was provided for participation.

### Primary and Secondary Outcomes

The primary outcome measure was patient-reported barriers to EPP access. Barriers to access were derived from the HINTS and secondarily classified as barriers of physical access, discomfort with technology, or patient preferences for nonelectronic provider communication (Multimedia Appendix 1) [11]. These categories were inductively coded by the research team after data collection. This classification is not validated, which we discuss as a limitation. Patients were instructed to mark preference for nonelectronic provider communication as a selection only if they perceived their preferences as barriers to using their portals. Importantly, "I do not have a patient portal account" was listed as an option on the original survey but was analyzed separately. The secondary outcome measure was patient-reported level of EPP use classified into 2 categories: routine use and nonroutine use. Per Maroney et al [17], level of use was characterized as routine if at least monthly use was indicated and as nonroutine if use a few times a year or less frequently was indicated, including those who did not have an EPP (Multimedia Appendix 1). Importantly, the level of EPP use included use for any clinic, not limited only to their orthopedic surgery care.

### Variables and Demographics

eHealth literacy was measured via patient responses to the eHEALS tool to determine its association with barriers to EPP use (Multimedia Appendix 2) [18]. This tool has been validated in the orthopedic outpatient setting, among others [18-20]. As in prior literature, a cumulative score of 25 or less indicated low eHealth literacy, and a score of 26 or greater indicated high eHealth literacy [13,21,22]. Neighborhood-level disadvantage was assessed using the ADI, which is calculated via publicly available census data in the domains of income, educational level, employment, and housing quality to assign numeric scores of societal disadvantage to particular geographical regions [23]. Higher scores indicate higher levels of societal disadvantage. Self-reported demographic data were additionally collected.

### Statistical Analysis

Survey data were analyzed using the Python statistical program (version 3.10.6; Python Software Foundation) [24-26]. Missing values were excluded pointwise across the relevant analyses given the low frequency. Numerical data presented as medians were reported with the IQR. The level of significance was set at $P$=.05.

Bivariate analyses were performed to examine the association between both patient-reported barriers and level of EPP use and demographic variables, the ADI, and eHealth literacy level. Three multivariable logistic regressions were conducted wherein barriers were treated as a binary outcome (1=barrier; 0=no barrier). Categorical variables were converted to binary dummy variables for regression. The main regression considered all barriers, whereas the 2 subsequent regressions examined only barriers of technology discomfort or preference-related barriers. Regression was not performed for physical access–related barriers due to outcome size of 13, which is discussed as a limitation. Variable selection for each model was determined via outcome size (at least 10 outcomes per covariate to avoid overfitting) and one-at-a-time sensitivity analyses (via examination of the McFadden pseudo-$R^2$) in a backward regression approach. To avoid collinearity, only covariates with a variance inflation factor of <5 were included together. Model fit was assessed using McFadden pseudo-$R^2$ values, with 0.2 considered excellent if not overfit (corroborated via df).

Demographic categories and regression reference levels were selected based on breakdowns and historical controls used in prior related literature [9,27,28]. Income was treated categorically, and per the analogous literature, we selected 3 levels representative of low, medium, and high income based on the median household income cutoffs for the zip code tied to the authors' institution.

## Results

### Participant Characteristics

A convenience sample of 339 eligible patients was approached, of whom 52 (15.3%) declined participation, leaving 287 (84.7%) for analysis. The median age of the study participants was 48.5 (IQR 35.0-64.2) years; 58.2% (167/287) of the study participants self-identified as non-Hispanic Black individuals, and 26.1% (75/287) identified as non-Hispanic White individuals. The median cumulative eHEALS score was 32 (IQR 27-35), with 21.3% (61/287) of the study participants having low eHealth literacy (eHEALS score of 25 or less). The median ADI was 53.0 (IQR 32.0-74.8).

Of the study participants, 63.1% (181/287) were routine users, and only 9.8% (28/287) did not have EPPs. One or more barriers to accessing their EPPs were reported by 43.2% (124/287) of all patients. Moreover, among the 90.2% (259/287) of patients who were enrolled in the EPP, 42.5% (110/259) still reported barriers to access or use of their portal. The remaining patient characteristics were compared by self-reported barriers to EPP access and self-reported EPP use (Table 1). Patients reporting one or more barriers had higher median age than patients who did not report barriers (57.0, IQR 42.2-71.0 years vs 43.0, IQR

29.2-57.0 years; *P*<.001; Table 1). A higher percentage of non-Hispanic Black patients (*P*=.04), retirees (*P*=.001), and patients who fell into the income bracket of US $30,000 or less (*P*=.03) reported barriers to access (Table 1). Patients who did not routinely use the EPP had a lower educational level than routine users (*P*=.005; Table 1).

**Table .** Descriptive characteristics and exploratory comparison of self-reported barriers to electronic patient portal (EPP) access and self-reported EPP use by patient characteristics.[a]

| Characteristic | Overall | No barriers | One or more barriers | *P* value | Routine use | Nonroutine use | *P* value |
|---|---|---|---|---|---|---|---|
| Subspecialty clinic, n/N (%) | | | | .07 | | | .77 |
| Hand | 187/287 (65.2) | 111/187 (59.4) | 76/187 (40.6) | | 116/187 (62.0) | 71/187 (38.0) | |
| Foot and ankle | 68/287 (23.7) | 40/68 (58.8) | 28/68 (41.2) | | 43/68 (63.2) | 25/68 (36.8) | |
| Trauma | 32/287 (11.1) | 12/32 (37.5) | 20/32 (62.5) | | 22/32 (68.8) | 10/32 (31.3) | |
| Age (years), median (IQR) | 48.5 (35.0-64.2) | 43.0 (29.2-57.0) | 57.0 (42.2-71.0) | <.001 | 47.0 (35.0-62.0) | 52.0 (37.0-67.0) | .28 |
| Race or ethnicity, n/N (%) | | | | .04 | | | .09 |
| Hispanic or Latino | 24/287 (8.4) | 16/24 (66.7) | 8/24 (33.3) | | 12/24 (50.0) | 12/24 (50.0) | |
| Non-Hispanic Black | 167/287 (58.2) | 83/167 (49.7) | 84/167 (50.3) | | 103/167 (61.7) | 64/167 (38.3) | |
| Non-Hispanic White | 75/287 (26.1) | 51/75 (68.0) | 24/75 (32.0) | | 48/75 (64.0) | 27/75 (36.0) | |
| Other identity or preferred not to answer | 21/287 (7.3) | 13/21 (61.9) | 8/21 (38.0) | | 18/21 (85.7) | 3/21 (14.3) | |
| Highest educational level attained, n/N (%) | | | | .07 | | | .005 |
| High school or lower | 82/284 (28.9) | 39/82 (47.6) | 43/82 (52.4) | | 41/82 (50.0) | 41/82 (50.0) | |
| Some college | 73/284 (25.7) | 40/73 (54.8) | 33/73 (45.2) | | 45/73 (61.6) | 28/73 (38.4) | |
| College or higher | 129/284 (45.4) | 82/129 (63.6) | 47/129 (36.4) | | 93/129 (72.1) | 36/129 (27.9) | |
| Annual income bracket (US $), n/N (%) | | | | .03 | | | .09 |
| ≤30,000 | 89/260 (34.2) | 42/89 (47.2) | 47/89 (52.8) | | 52/89 (58.4) | 37/89 (41.6) | |
| 30,001-50,000 | 48/260 (18.5) | 26/48 (54.2) | 22/48 (45.8) | | 30/48 (62.5) | 18/48 (37.5) | |
| >50,000 | 123/260 (47.3) | 80/123 (65.0) | 43/123 (35.0) | | 89/123 (72.4) | 34/123 (27.6) | |
| Current employment status, n/N (%) | | | | .001 | | | .19 |
| Employed | 160/285 (56.1) | 105/160 (65.6) | 55/160 (34.4) | | 109/160 (68.1) | 51/160 (31.9) | |
| Unemployed or on disability | 66/285 (23.2) | 35/66 (53.0) | 31/66 (47.0) | | 38/66 (57.6) | 28/66 (42.4) | |
| Retired | 59/285 (20.7) | 22/59 (37.3) | 37/59 (62.7) | | 34/59 (57.6) | 25/59 (42.4) | |

[a]Percentages may not add up to 100 due to rounding.

## Exploratory Analysis of eHealth and ADI as Potential Barriers to EPP Access and Use

In the analysis of eHealth literacy, patients who reported barriers had a lower median eHEALS score than patients who did not report barriers (29.0, IQR 22.0-32.2 vs 32.0, IQR 30.0-38.0; *P*<.001; Table 2). Conversely, a higher percentage of patients with low eHealth literacy compared to high eHealth literacy reported barriers to using their EPPs (45/61, 74% vs 79/224, 35%, respectively; *P*<.001; Table 2). Patients who reported barriers also had higher median national ADI than patients who did not report barriers (55.5, IQR 37.5-78.2 vs 51.0, IQR 32.0-70.0; *P*=.06), and a higher percentage of patients from the most deprived ADI quartile also reported barriers compared to patients from the least deprived ADI quartiles (most deprived ADI quartile [76-100], 37/70, 53% vs second most deprived ADI quartile [51-75], 35/87, 40% vs second least deprived ADI quartile [26-50], 29/70, 41% vs least deprived ADI quartile [1-25], 19/55, 35%; *P*=.19); however, these results did not reach statistical significance (Table 2).

**Table .** Comparison of self-reported barriers to electronic patient portal access by eHealth literacy level and the Area Deprivation Index (ADI).[a]

| | Overall | No barriers | Any barriers | *P values* |
|---|---|---|---|---|
| eHEALS[b] score, median (IQR) | 32.0 (27.0-35.0) | 32.0 (30.0-38.0) | 29.0 (22.0-32.2) | <.001 |
| eHealth literacy level (eHEALS score), n/N (%) | | | | <.001 |
| High eHealth literacy | 224/285 (78.6) | 145/224 (64.7) | 79/224 (35.3) | |
| Low eHealth literacy | 61/285 (21.4) | 16/61 (26.2) | 45/61 (73.8) | |
| National ADI, median (IQR) | 53.0 (32.0-74.8) | 51.0 (32.0-70.0) | 55.5 (37.5-78.2) | .06 |
| National ADI quartile, n/N (%) | | | | .19 |
| Least deprived ADI quartile (1-25) | 55/282 (19.5) | 36/55 (65.4) | 19/55 (34.5) | |
| Second least deprived ADI quartile (26-50) | 70/282 (24.8) | 41/70 (58.6) | 29/70 (41.4) | |
| Second most deprived ADI quartile (51-75) | 87/282 (30.9) | 52/87 (59.8) | 35/87 (40.2) | |
| Most deprived ADI quartile (76-100) | 70/282 (24.8) | 33/70 (47.1) | 37/70 (52.9) | |

[a]Percentages may not add up to 100 due to rounding and missing values.

[b]eHEALS: eHealth Literacy Scale.

Most of the patient-reported barriers were related to technology discomfort (55/124, 44.4%) or preference (78/124, 62.9%) rather than physical access (13/124, 10.5%). However, the group reporting physical access barriers had the lowest levels of eHealth literacy (median eHEALS score 17.0 vs 27.0 vs 27.0, respectively) and highest ADI (median 73.0 vs 53.5 vs 58.0, respectively) compared to groups reporting barriers related to technology discomfort or preference (Table 3). Patients from the most deprived ADI quartile had higher percentages of barriers (Table 3). Bivariate analysis was not performed as barrier type was nonexclusive.

**Table .** Exploratory analysis of barrier type by eHealth literacy level and the Area Deprivation Index (ADI).[a]

| | No barriers | Physical access barriers | Technology discomfort barriers | Preference barriers |
|---|---|---|---|---|
| eHEALS[b] score, median (IQR) | 32.0 (30.0-38.0) | 17.0 (12.0-24.0) | 27.0 (16.0-32.0) | 27.0 (20.0-32.0) |
| eHealth literacy level (eHEALS), n/N (%) | | | | |
| High eHealth literacy | 145/224 (64.7) | 2/224 (0.9) | 30/224 (13.4) | 44/224 (19.6) |
| Low eHealth literacy | 16/61 (26.2) | 11/61 (18.0) | 25/61 (41.0) | 34/61 (55.7) |
| National ADI, median (IQR) | 51.0 (32.0-70.0) | 73.0 (41.0-92.0) | 53.5 (31.2-76.0) | 58.0 (38.8-83.8) |
| National ADI quartile, n/N (%) | | | | |
| Least deprived ADI quartile (1-25) | 36/55 (65.5) | 0/55 (0.0) | 10/55 (18.2) | 10/55 (18.2) |
| Second least deprived ADI quartile (26-50) | 41/70 (58.6) | 4/70 (5.7) | 13/70 (18.6) | 19/70 (27.1) |
| Second most deprived ADI quartile (51-75) | 52/87 (59.8) | 3/87 (3.4) | 17/87 (19.5) | 15/87 (17.2) |
| Most deprived ADI quartile (76-100) | 33/70 (47.1) | 6/70 (8.6) | 14/70 (20.0) | 30/70 (42.9) |

[a]Percentages may not add up to 100 due to rounding and missing values.

[b]eHEALS: eHealth Literacy Scale.

## Regression Analysis of Barriers to EPP Use

In the overall regression including demographic variables, the ADI, and eHealth literacy level, only low eHealth literacy level (adjusted odds ratio [AOR] 1.32, 95% CI 1.13-1.54; *P*<.001) and older age (AOR 1.007, 95% CI 1.003-1.009; *P*<.001) predicted barriers to EPP access (Table 4). Similarly, only low eHealth literacy level and age were associated with predicting

a technology discomfort–related barrier (low eHealth literacy, AOR 1.25, 95% CI 1.11-1.40; *P*<.001; age, AOR 1.004, 95% CI 1.002-1.007; *P*<.001; Table 4) or a preference-related barrier (low eHealth literacy, AOR 1.33, 95% CI 1.17-1.51; *P*<.001; age, AOR 1.004, 95% CI 1.00-1.01; *P*=.01; Table 4). The ADI was not associated with predicting overall barriers (*P*=.59), preference-related barriers (*P*=.35), or technology discomfort–related barriers (*P*=.76; Table 4).

**Table .** Regression of patient characteristics associated with self-reporting at least one barrier to electronic patient portal use (any barrier type, preference barrier, or technology discomfort barrier).

| Characteristic | Regression 1: any barrier type[a] | | Regression 2: preference barriers[b] | | Regression 3: technology discomfort barriers[c] | |
|---|---|---|---|---|---|---|
| | AOR[d] (95% CI) | *P* value | AOR (95% CI) | *P* value | AOR (95% CI) | *P* value |
| Age | 1.007 (1.003-1.009) | <.001 | 1.004 (1.00-1.01) | .01 | 1.004 (1.002-1.007) | <.001 |
| Race or ethnicity | | | | | | |
| Hispanic or Latino | 0.99 (0.78-1.25) | .92 | 1.07 (0.88-1.31) | .49 | 1.04 (0.86-1.24) | .71 |
| Non-Hispanic Black | 1.09 (0.94-1.27) | .26 | 1.09 (0.95-1.24) | .21 | 1.04 (0.92-1.17) | .52 |
| Non-Hispanic White | Reference | —[e] | Reference | — | Reference | — |
| Other identity or preferred not to answer | 1.02 (0.79-1.32) | .88 | 1.05 (0.85-1.30) | .63 | 0.95 (0.78-1.15) | .56 |
| National ADI[f] | 1.00 (1.00-1.00) | .59 | 1.00 (1.00-1.00) | .35 | 1.00 (1.00-1.00) | .76 |
| eHealth literacy level (eHEALS[g]) | | | | | | |
| High eHealth literacy | Reference | — | Reference | — | Reference | — |
| Low eHealth literacy | 1.32 (1.13-1.54) | <.001 | 1.33 (1.17-1.51) | <.001 | 1.25 (1.11-1.40) | <.001 |
| Subspecialty clinic | | | | | | |
| Hand | Reference | — | Reference | — | — | — |
| Foot and ankle | 1.01 (0.88-1.16) | .91 | 1.02 (0.91-1.15) | .73 | — | — |
| Trauma | 1.18 (0.99-1.41) | .07 | 1.17 (1.00-1.37) | .05 | — | — |
| Highest educational level attained | | | | | | |
| High school or lower | 1.06 (0.90-1.25) | .65 | — | — | — | — |
| Some college | 0.96 (0.82-1.13) | .51 | — | — | — | — |
| College or higher | Reference | — | — | — | — | — |
| Annual income bracket (US $) | | | | | | |
| ≤30,000 | 1.11 (0.94-1.30) | .22 | — | — | — | — |
| 30,001-50,000 | 1.05 (0.89-1.24) | .59 | — | — | — | — |
| >50,000 | Reference | — | — | — | — | — |

[a]Regression model 1: 88.5% (254/287) of the patients were included after participants with missing values were excluded (outcome size: 124/254, 48.8% reported any barrier; df=12; pseudo-$R^2$=0.21).

[b]Regression model 2: 96.5% (277/287) of the patients were included after participants with missing values were excluded (outcome size: 78/277, 28.2% reported preference barriers; df=8; pseudo-$R^2$=0.16).

[c]Regression model 3: 96.5% (277/287) of the patients were included after participants with missing values were excluded (outcome size: 55/277, 19.9% reported technology discomfort barriers; df=6; pseudo-$R^2$=0.13).

[d]AOR: adjusted odds ratio.

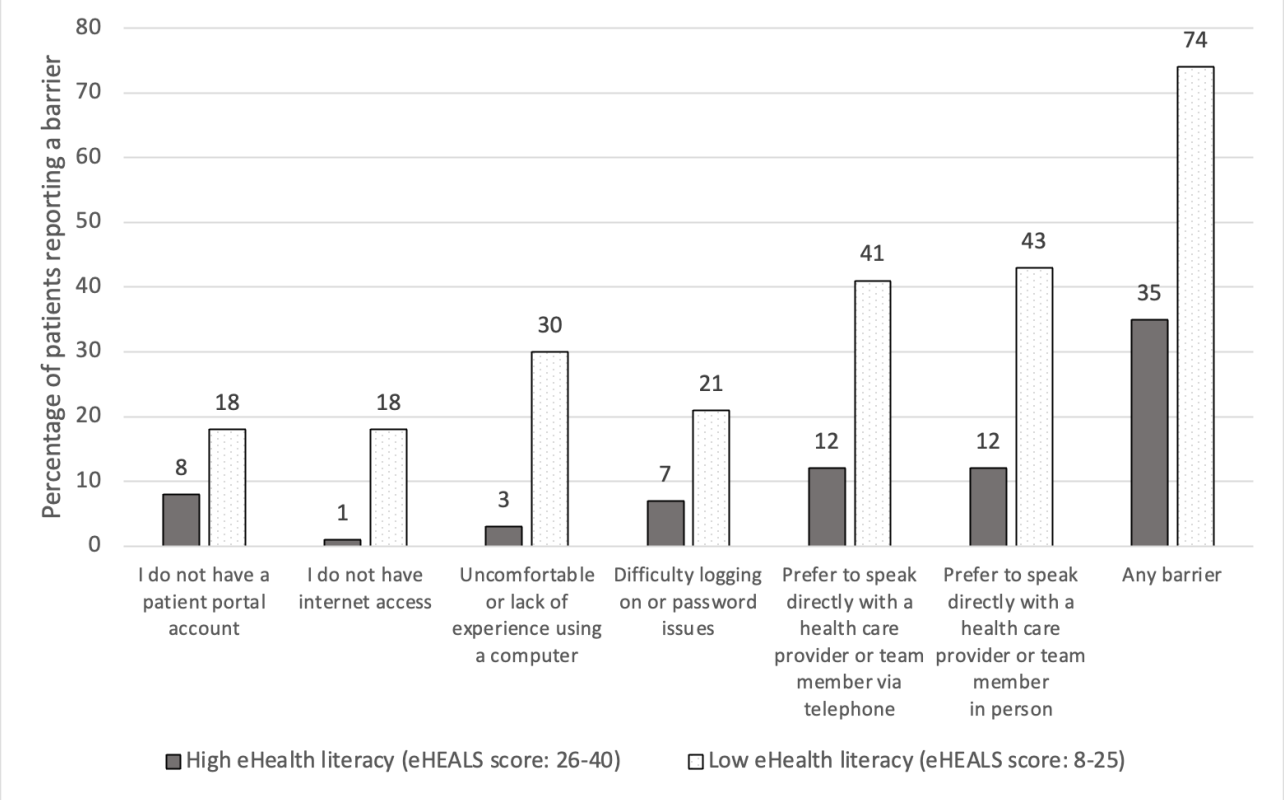
[e]Not applicable.

[f]ADI: Area Deprivation Index.

[g]eHEALS: eHealth Literacy Scale.

The most prevalent patient-reported barrier to EPP access was a preference to speak directly to the health care provider or team member in person or via telephone (52/287, 18.1%). A lack of comfort with a computer was cited as a barrier to EPP access by 29.5% (18/61) of patients with low eHealth literacy (Figure 1). Additionally, 17.8% (51/287) of study participants reported poor or fair ability to use a computer, tablet, or smartphone to find information that they needed on the internet. One or more barriers to EPP use were indicated by 80.4% (41/51) of patients with poor or fair self-ratings, compared to 35.3% (83/235) of patients with good, very good, or excellent self-ratings ($P$<.001).

**Figure 1.** Frequency and percentages of patient self-reported barriers to electronic patient portal (EPP) access stratified by eHealth literacy level. The percentages of patients who reported a particular barrier are reported over the denominator of patients with either high eHealth literacy scores (n=224) or low eHealth literacy scores (n=61). Percentages do not necessarily add up to 100 across barrier categories as patients could indicate more than one barrier. The barriers "I do not need a patient portal account," "I have multiple patient portals," and "I have privacy concerns" are not depicted separately due to lower frequency of observations.




## Discussion

### Principal Findings

Equitable implementation of digitized health tools relies on efforts from clinicians, researchers, and policymakers alike. This study assessed patient-reported barriers to use within an expanded framework of individual- and neighborhood-level factors. Low eHealth literacy level was the most significant determinant of overall barriers to EPP use for orthopedic surgery, followed by older age, as compared to other demographic factors and measures of neighborhood-level disadvantage. Contrary to expectation, neighborhood-level disadvantage as measured via the ADI had no mediating effect on barriers after adjusting for eHealth literacy level. Patients with physical access barriers did have appreciably higher ADIs; however, few patients reported physical access barriers overall. These findings build on prior work that showed that older age, among other patient demographics, was associated with reduced EPP enrollment in orthopedic surgery [3,4,8]. This also builds on prior work in the hospital medicine setting that showed that lower eHealth literacy correlated with decreased awareness and use and less favorable attitudes toward use of EPPs [13].

Barriers of physical access, such as lack of internet access, were infrequent compared to barriers related to preference or discomfort with technology, including lack of experience using a computer or difficulty logging in. This finding contrasts with previous work that found internet access to be a significant determinant of portal use for orthopedic surgery [6]. Additionally, among the 17.8% (51/287) of the participants with lower self-ratings of ability to use a computer, tablet, or smartphone to find information that they needed on the internet, 80.4% (41/51) indicated barriers to accessing their EPPs. The final rule of the 21st Century Cures Act alleviated many barriers related to physical access (eg, computer access or broadband coverage) by requiring an interoperability standard that any EPP programming interface be compatible with smartphone apps [29]. However, this legislation does not address barriers experienced by patients who may technically have the physical and digital tools to access their EPPs but not the self-efficacy or the skill sets to effectively use these tools.

These findings may support a second digital divide being dependent on disparities in skill sets rather than physical access [13,30]. With regard to EPPs, this is clinically significant as most institutions introduce EPPs via email with time-sensitive

links and sign-up instructions. This may be ineffective at best in promoting EPP adoption or postenrollment use in populations who may have internet access yet lack the internet experience and skills to navigate setting up a digital account with time-pressure activation codes for protected health information. Moreover, the finding that over a quarter of all patients (52/287, 18.1%) perceived their preference to speak with a health care provider in person or via telephone as an actual barrier to using the EPP suggests that patients may view the EPP as a substitute for health care provider communication missing the personal element rather than as an adjunct to improve communication and transparency, as it was intended. This finding also suggests that it may be important to document patient preference for EPP versus non-EPP communication even if a patient does have an EPP as patients with activated EPPs may not be active users.

Importantly, while most patients were enrolled in the EPP, a significant portion of enrolled patients reported barriers to access and use of their portal. This is significant as prior studies of EPP use for orthopedic surgery have either only assessed EPP activation status as a surrogate for use and without assessing barriers to use [3,4,8] or qualitatively analyzed optional comments regarding nonuse in a small fraction of the study cohort (38 of 150 patients) [6]. These findings are clinically relevant as prior efforts to reduce disparities in EPP use have also focused on enrollment [9]. However, simply enrolling patients does little to address the underlying barriers of certain patients to using their EPPs. Enrollment will not address individual-level factors such as eHealth literacy, which may constitute a larger underlying barrier to sustained EPP use after enrollment (eg, patients' technological capabilities and skills rather than physical access to the technology itself).

Additionally, this study substantiates that older adults are a vulnerable population that may be left behind in a digitized health system. This is particularly critical to perioperative care in orthopedic surgery. Older patients may have more complex discharge needs, including perioperative medication changes and rehabilitation requiring close postoperative communications, and addressing them within the EPP may be ineffective. Older patients have intersecting factors that impede their access. While age and eHealth literacy were noncollinear in this study, age has well-known associations with traditional health and eHealth literacy [31]. Addressing deficiencies in these skill sets may improve lower levels of self-efficacy to adopt and use EPPs previously reported in older individuals [32]. This group may benefit from a proactive staff-level intervention that supports an in-person EPP enrollment option followed by an initial lesson on how to use the EPP, which Bhashyam et al [33] previously showed may be beneficial to improving postoperative follow-up. Notably, while older patients may have an elevated sense of caution in using online platforms due to counseling from groups such as the American Association of Retired Persons, patients infrequently noted privacy concerns as a barrier to EPP use in this study.

Patients with physical access barriers also had appreciably higher ADIs despite this analysis not meeting statistical significance. These patients may also benefit from routine touchpoints with staff outside of the EPP. Ensuring effective follow-up is especially important in these patients as, in addition to worse ADIs predicting worse comorbid chronic disease outcomes, simply living in a disadvantaged neighborhood confers similar readmission risk as having a chronic lung disease and higher risk than having a chronic condition such as diabetes [34]. At an informatics design level, a widget within the electronic medical record could be implemented to automatically yield an address-generated ADI analogously to how BMI is automatically calculated based on a patient's weight and height as this may generate similarly important contextual information to a patient's overall health, especially in a perioperative context.

## Limitations

First, this was a single-institution study conducted across several orthopedic surgery subspecialty clinics not including adult reconstruction. Hence, the results may not be generalizable across other settings. Moreover, while eHEALS is a commonly used, validated screener for eHealth literacy in outpatient settings [18-20], it has not been validated in our specific patient population. Additionally, self-reporting via a survey may be limited by response bias; however, we felt that this method of examination was critical to include the patient perspective. Importantly, this study did not include non–English-speaking patients, who may experience additional barriers to EPP access; however, this study did include the perspectives of patients with limited reading and writing skills.

Additionally, level of use was dichotomized as routine and nonroutine, similar to the study by Maroney et al [17], without an option for "as needed," which assumes regular use. The National Cancer Institute's 2020 HINTS showed that only 40% of those with EPPs used it every year: 65% felt that they did not need to use it every year [11]. An additional checkbox option for "use as needed" may better capture this nuance; however, this could introduce indeterminate subjectivity.

Notably, the sample size may have been underpowered to detect small effect sizes in demographic differences and in novel outcomes such as the ADI. The sample size calculation was predicated on the 2020 HINTS, which analyzed barriers to EPP enrollment and did not include patients with activated EPPs who still experienced barriers (259/287, 90.2% of our study population) [11]. Moreover, no relevant existing literature has assessed the ADI. Finally, the secondary barrier categories were inductively coded by our research team after data collection based on natural groupings in which we were interested. This rendered our initial sample size calculation insufficient to perform a secondary regression analysis for the access-related barrier category, which had a small outcome size. Additionally, this classification was not validated, which may introduce bias but, importantly, allowed for discovery of new insights that may not have been generated through precoding.

## Conclusions

Routine use of EPPs for online scheduling, patient communication, and telehealth continues to be a critical aspect of care. It is necessary to understand existing disparities in barriers to EPP access to not only improve access to care for all patients but also to continue building patients' toolbox and self-efficacy to take on active roles in their care. Future research

should establish whether interventions, education, and improved eHealth literacy may overcome these barriers.

## Acknowledgments

The authors would like to thank Jennifer Wolf, MD, and Megan Conti Mica, MD, for helping facilitate data collection within their outpatient clinics. The authors attest that there was no use of generative artificial intelligence technology in the generation of text, figures, or other informational content of this manuscript.

## Funding

The authors did not receive support from any organization for the submitted work. No funding was received to assist with the preparation of this manuscript. No funding was received for conducting this study.

## Data Availability

The datasets generated or analyzed during this study are not publicly available due to institutional review board constraints but are available from the corresponding author on reasonable request.

## Authors' Contributions

ALL, NL, KH, JS, and JGS contributed equally to the design, data collection, and writing of the manuscript.

Each author certifies that there are no funding or commercial associations (consultancies, stock ownership, equity interest, patent/licensing arrangements, etc) that might pose a conflict of interest in connection with the submitted article related to the author or any immediate family members.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Summary of survey questions regarding patient portal.
[DOCX File, 15 KB - periop_v9i1e72035_app1.docx ]

Multimedia Appendix 2
Summary of eHealth Literacy Scale items.
[DOCX File, 15 KB - periop_v9i1e72035_app2.docx ]

## References

1. Dendere R, Slade C, Burton-Jones A, Sullivan C, Staib A, Janda M. Patient portals facilitating engagement with inpatient electronic medical records: a systematic review. J Med Internet Res 2019 Apr 11;21(4):e12779. [doi: 10.2196/12779] [Medline: 30973347]
2. Davis Giardina T, Menon S, Parrish DE, Sittig DF, Singh H. Patient access to medical records and healthcare outcomes: a systematic review. J Am Med Inform Assoc 2014;21(4):737-741. [doi: 10.1136/amiajnl-2013-002239] [Medline: 24154835]
3. Varady NH, d'Amonville S, Chen AF. Electronic patient portal use in orthopaedic surgery is associated with disparities, improved satisfaction, and lower no-show rates. J Bone Joint Surg Am 2020 Aug 5;102(15):1336-1343. [doi: 10.2106/JBJS.19.01080] [Medline: 32769600]
4. Vallurupalli N, Lawrence KW, Habibi AA, Bosco JA, Lajam CM. Socioeconomic disparities in online patient portal utilization among total knee arthroplasty recipients. J Arthroplasty 2024 Sep;39(9S2):S95-S99. [doi: 10.1016/j.arth.2024.04.040] [Medline: 38670173]
5. Plate JF, Ryan SP, Bergen MA, Hong CS, Attarian DE, Seyler TM. Utilization of an electronic patient portal following total joint arthroplasty does not decrease readmissions. J Arthroplasty 2019 Feb;34(2):211-214. [doi: 10.1016/j.arth.2018.11.002] [Medline: 30497899]
6. Kolodychuk NL, Wong M, Chimento G, Adams T, Gastanaduy M, Waddell BS. Web-based patient portal access in an orthopedic adult reconstruction patient population. Arthroplast Today 2019 Feb 22;5(1):83-87. [doi: 10.1016/j.artd.2019.01.004] [Medline: 31020029]
7. Fatehi A, Gonzalez A, Ring DC, Queralt M. Psychosocial factors are associated with electronic portal registration. Clin Orthop Relat Res 2020 Sep;478(9):2079-2084. [doi: 10.1097/CORR.0000000000001278] [Medline: 32332246]
8. Bell K, Warnick E, Nicholson K, et al. Patient adoption and utilization of a web-based and mobile-based portal for collecting outcomes after elective orthopedic surgery. Am J Med Qual 2018;33(6):649-656. [doi: 10.1177/1062860618765083] [Medline: 29562769]

9. Richwine C, Johnson C, Patel V. Disparities in patient portal access and the role of providers in encouraging access and use. J Am Med Inform Assoc 2023 Jan 18;30(2):308-317. [doi: 10.1093/jamia/ocac227] [Medline: 36451262]
10. Anthony DL, Campos-Castillo C, Lim PS. Who isn't using patient portals and why? Evidence and implications from a national sample of US adults. Health Aff (Millwood) 2018 Dec;37(12):1948-1954. [doi: 10.1377/hlthaff.2018.05117] [Medline: 30633673]
11. Johnson C, Richwine C, Patel V. Individuals' access and use of patient portals and smartphone health apps, 2020. In: ASTP Health IT Data Brief [Internet]: Office of the Assistant Secretary for Technology Policy; 2021.
12. Nishii A, Campos-Castillo C, Anthony D. Disparities in patient portal access by US adults before and during the COVID-19 pandemic. JAMIA Open 2022 Dec;5(4):ooac104. [doi: 10.1093/jamiaopen/ooac104] [Medline: 36540762]
13. Deshpande N, Arora VM, Vollbrecht H, Meltzer DO, Press V. eHealth literacy and patient portal use and attitudes: cross-sectional observational study. JMIR Hum Factors 2023 Jan 27;10:e40105. [doi: 10.2196/40105] [Medline: 36705947]
14. Lyles CR, Nelson EC, Frampton S, Dykes PC, Cemballi AG, Sarkar U. Using electronic health record portals to improve patient engagement: research priorities and best practices. Ann Intern Med 2020 Jun 2;172(11 Suppl):S123-S129. [doi: 10.7326/M19-0876] [Medline: 32479176]
15. Lyles CR, Nguyen OK, Khoong EC, Aguilera A, Sarkar U. Multilevel determinants of digital health equity: a literature synthesis to advance the field. Annu Rev Public Health 2023 Apr 3;44:383-405. [doi: 10.1146/annurev-publhealth-071521-023913] [Medline: 36525960]
16. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208. [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
17. Maroney K, Curtis LM, Opsasnick L, et al. eHealth literacy and web-based patient portal usage among kidney and liver transplant recipients. Clin Transplant 2021 Feb;35(2):e14184. [doi: 10.1111/ctr.14184] [Medline: 33278846]
18. Norman CD, Skinner HA. eHEALS: the eHealth Literacy Scale. J Med Internet Res 2006 Nov 14;8(4):e27. [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
19. Lee J, Lee EH, Chae D. eHealth literacy instruments: systematic review of measurement properties. J Med Internet Res 2021 Nov 15;23(11):e30644. [doi: 10.2196/30644] [Medline: 34779781]
20. Cherid C, Baghdadli A, Wall M, et al. Current level of technology use, health and eHealth literacy in older Canadians with a recent fracture-a survey in orthopedic clinics. Osteoporos Int 2020 Jul;31(7):1333-1340. [doi: 10.1007/s00198-020-05359-3] [Medline: 32112118]
21. Richtering SS, Hyun K, Neubeck L, et al. eHealth literacy: predictors in a population with moderate-to-high cardiovascular risk. JMIR Hum Factors 2017 Jan 27;4(1):e4. [doi: 10.2196/humanfactors.6217] [Medline: 28130203]
22. Vollbrecht H, Arora VM, Otero S, Carey KA, Meltzer DO, Press VG. Measuring eHealth literacy in urban hospitalized patients: implications for the post-COVID world. J Gen Intern Med 2021 Jan;36(1):251-253. [doi: 10.1007/s11606-020-06309-9] [Medline: 33078294]
23. Kind AJ, Buckingham WR. Making neighborhood-disadvantage metrics accessible - the Neighborhood Atlas. N Engl J Med 2018 Jun 28;378(26):2456-2458. [doi: 10.1056/NEJMp1802313] [Medline: 29949490]
24. The Python language reference. Python. URL: https://docs.python.org/3/reference/ [accessed 2022-08-31]
25. Virtanen P, Gommers R, Oliphant TE, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 2020 Mar;17(3):261-272. [doi: 10.1038/s41592-019-0686-2] [Medline: 32015543]
26. Seabold S, Perktold J. Statsmodels: econometric and statistical modeling with Python. Presented at: Proceedings of the 9th Python in Science Conference; Jun 28 to Jul 3, 2010. [doi: 10.25080/Majora-92bf1922-011]
27. El-Toukhy S, Méndez A, Collins S, Pérez-Stable EJ. Barriers to patient portal access and use: evidence from the health information national trends survey. J Am Board Fam Med 2020;33(6):953-968. [doi: 10.3122/jabfm.2020.06.190402] [Medline: 33219074]
28. Lockwood MB, Dunn-Lopez K, Pauls H, Burke L, Shah SD, Saunders MA. If you build it, they may not come: modifiable barriers to patient portal use among pre- and post-kidney transplant patients. JAMIA Open 2018 Jul 10;1(2):255-264. [doi: 10.1093/jamiaopen/ooy024] [Medline: 31984337]
29. ONC's Cures Act Final Rule. HealthIT.gov. URL: https://www.healthit.gov/topic/oncs-cures-act-final-rule [accessed 2025-12-24]
30. van Deursen A, van Dijk J. Internet skills and the digital divide. New Media Soc 2010;13(6):893-911. [doi: 10.1177/1461444810386774]
31. Chesser AK, Keene Woods N, Smothers K, Rogers N. Health literacy and older adults: a systematic review. Gerontol Geriatr Med 2016 Mar 15;2:2333721416630492. [doi: 10.1177/2333721416630492] [Medline: 28138488]
32. Price-Haywood EG, Harden-Barrios J, Ulep R, Luo Q. eHealth literacy: patient engagement in identifying strategies to encourage use of patient portals among older adults. Popul Health Manag 2017 Dec;20(6):486-494. [doi: 10.1089/pop.2016.0164] [Medline: 28384076]
33. Bhashyam AR, Bansal M, McGovern MM, van der Vliet QMJ, Heng M. In-hospital enrollment into an electronic patient portal results in improved follow-up after orthopedic surgery: cluster randomized controlled trial. JMIR Perioper Med 2022 Aug 11;5(1):e37148. [doi: 10.2196/37148] [Medline: 35969452]

34. Kind AJ, Jencks S, Brock J, et al. Neighborhood socioeconomic disadvantage and 30-day rehospitalization: a retrospective cohort study. Ann Intern Med 2014 Dec 2;161(11):765-774. [doi: 10.7326/M13-2946] [Medline: 25437404]

## Abbreviations

**ADI:** Area Deprivation Index
**AOR:** adjusted odds ratio
**eHEALS:** eHealth Literacy Scale
**EPP:** electronic patient portal
**HINTS:** Health Information National Trends Survey
**HIPAA:** Health Insurance Portability and Accountability Act
**REDCap:** Research Electronic Data Capture
**STROBE:** Strengthening the Reporting of Observational Studies in Epidemiology

*Edited by N Rohatgi; submitted 03.Feb.2025; peer-reviewed by A Tsehay, M Rincon, X Ding; revised version received 06.Dec.2025; accepted 09.Dec.2025; published 07.Jan.2026.*

*Please cite as:*
*Litvak AL, Lin N, Hynes K, Strelzow J, Stepan JG*
*Assessing the Effects of eHealth Literacy and the Area Deprivation Index on Barriers to Electronic Patient Portal Use for Orthopedic Surgery: Cross-Sectional Observational Study*
*JMIR Perioper Med 2026;9:e72035*
*URL: https://periop.jmir.org/2026/1/e72035*
*doi:10.2196/72035*

# Benchmark Development for Fundamental Arthroscopic Skills Using a Simulation-Based Training Program: Observational Study

Eric Davis[1], MBA; Brianna Caraet[2], MD; Robert Pedowitz[3], MD, PhD; Gregg Nicandri[1], MD

[1]Department of Orthopaedic Surgery, University of Rochester, 601 Elmwood Avenue, Rochester, NY, United States

[2]Kaiser Permanente Orthopedic Sports Medicine, San Diego, CA, United States

[3]Department of Orthopaedic Surgery, University of California, Los Angeles, CA, United States

**Corresponding Author:**
Eric Davis, MBA
Department of Orthopaedic Surgery, University of Rochester, 601 Elmwood Avenue, Rochester, NY, United States

## Abstract

**Background:** Surgical education has shifted from the traditional Halstedian apprenticeship model toward incorporating simulation due to work-hour restrictions, increasing case complexity, and economic and liability pressures. Building on the success of the Fundamentals of Laparoscopic Surgery program for general surgery, the Fundamentals of Arthroscopic Surgery Training (FAST) program was developed to establish proficiency benchmarks for orthopedic trainees in basic arthroscopic skills.

**Objective:** We aimed to establish benchmarks for 5 FAST workstation modules.

**Methods:** Sports medicine fellowship–trained faculty members were given instructions on the modules and 2 minutes of practice time, and they then performed each task 3 times with both their dominant and nondominant hand. For each module, mean faculty performance was used to establish an efficiency benchmark (time) and precision benchmark (errors).

**Results:** The Probing module should be completed in less than 95 seconds with no errors. The Ring Transfer module should be completed in less than 134 seconds with no more than 1 error. The Maze module should be completed in less than 99 seconds with no errors. The Meniscectomy module should be completed in less than 68 seconds with no more than 1 error. Lastly, the Suture Passing module should be completed in less than 195 seconds with no more than 1 error.

**Conclusions:** The FAST workstation can be used as a proficiency-based learning tool for residents to safely and effectively develop arthroscopic skills outside of the operating room. These benchmarks were established via a method previously validated in surgical simulation and balance precision and efficiency for skills that are considered generalizable and transferable to arthroscopic surgeries.

***(JMIR Perioper Med 2026;9:e82723)*** doi:10.2196/82723

**KEYWORDS**

surgical education; orthopedic surgery; arthroscopy; simulation training; skills assessment; resident education

## Introduction

Historically, surgical education has been based upon the Halstedian apprenticeship model, in which surgical trainees primarily acquire technical skills in the operating room while assisting and later performing procedures on live patients. In recent years, there has been a paradigm shift away from this approach in orthopedic surgery [1]. This can be attributed to multiple factors, including but not limited to resident work-hour restrictions, increased case complexity and variety, the risk of liability, and social and economic pressure. Simulation in surgical education has been suggested as a potential avenue to fill the gap in training, allowing trainees opportunities to obtain valuable experience in a low-risk, more cost-effective environment. In concordance with this, the Accreditation Council for Graduate Medical Education now requires that residency programs include surgical skills training in their curricula; this mandate has also been echoed by the American Board of Orthopaedic Surgery and the Residency Review Committee.

Simulation as an educational tool has been validated extensively in the field of general surgery. The Fundamentals of Laparoscopic Surgery program, developed in 1997 by the Society of American Gastrointestinal and Endoscopic Surgeons, combines cognitive and physical skills components shown to improve a given student's laparoscopic abilities over the course of the curriculum [2]. The program has been studied rigorously and validated since its release [3-9]. Currently, a passing score is required by the American Board of Surgery as prerequisite to sit for the certification exam. This has been well received by trainees, surgeons, and the public. As a result of this successful program in general surgery, the field of orthopedics has explored the development of a similar program.

In 2011, the Arthroscopy Association of North America (AANA), the American Board of Orthopaedic Surgery, and the American Academy of Orthopaedic Surgeons came together in a joint effort to establish the Fundamentals of Arthroscopic Surgery Training (FAST) program. The program is based on the tenet that fundamental surgical skills are best obtained sequentially, and that proficiency can be established through the successful completion of a basic skills curriculum. A curriculum and simulator have been developed for the FAST program and proficiency benchmarks for knot tying have been established.

The aim of this study was to develop time- and error-based proficiency benchmarks for 5 fundamental arthroscopic skills modules within a simulation-based training program.

## Methods

### Overview

The AANA offered a voluntary skills training course for residents and faculty at the Orthopaedic Learning Center in Chicago, Illinois, on December 7 to 9, 2017, and January 18 to 20, 2018. During downtime in the course, participants were recruited and provided with an information sheet outlining participation in this study. Two separate cohorts of 20 faculty each completed orthopedic simulations to improve their skill. For each module, the participant was given a list of written instructions and allowed to practice for a period of 2 minutes. Then, each participant performed the task 3 times with their dominant hand and 3 times with their nondominant hand. The order of dominant/nondominant completion was randomized so that exactly half of the faculty started with their nondominant hand. The time it took each participant to complete the module as well as the number of errors committed was recorded. Cohort 1 completed the Probing, Ring Transfer, and Maze Navigation modules. Cohort 2 completed the Meniscectomy and Suture Passing modules.

All modules were completed using camera-based visualization rather than live arthroscopes. Depending on the module, visualization was provided using either a USB camera system or a stationary camera, consistent with the simulation equipment available in the skills laboratory. This approach was used to standardize task execution across participants and to support reproducibility in a controlled training environment.

Five simulation-based arthroscopic skills modules were evaluated, including Probing, Ring Transfer, Maze Navigation, Partial Meniscectomy, and Suture Passage. Each module assessed fundamental psychomotor skills relevant to arthroscopy, with efficiency measured by task completion time and precision measured by predefined error criteria. Detailed descriptions of each module are provided in Multimedia Appendix 1.

Descriptive statistics were used for demographic variables, experience, handedness, efficiency (time) score, and precision (error) score. Mean performance time for all participants was determined. Any value lying more than 2 SDs from the mean was excluded. Exclusion of values greater than 2 SDs from the mean was performed to minimize the influence of extreme outliers that may reflect momentary lapses, equipment issues, or deviations from standardized task execution rather than true expert performance. This approach has been used in prior simulation-based benchmarking studies to derive stable proficiency thresholds that reflect typical expert performance rather than maximal or anomalous values. Outlier exclusion affected a small proportion of observations and did not meaningfully alter the relative ordering of module performance or the final benchmark thresholds. The trimmed mean was used as the basis for determining the proficiency level benchmark for each task. Efficiency benchmarks were set at the minimum threshold of the mean time of the expert cohort. Precision benchmarks were set at the mean number of errors committed by participants rounded to the nearest whole number. The benchmark was established as the mean performance when data from both hands were aggregated. Aggregation of dominant and nondominant hand performance was chosen a priori to reflect the ambidextrous demands of arthroscopic surgery, in which instrument laterality varies by procedure and portal placement. Use of the mean performance across both hands was intended to establish benchmarks that promote bilateral skill development while remaining feasible and generalizable for trainee progression.

### Ethical Considerations

This study was reviewed by the University of Rochester Research Subjects Review Board and was granted exempt status with a waiver of documentation of informed consent. All participants were provided with an information sheet outlining the purpose, procedures, and voluntary nature of the study prior to participation. All data collected were deidentified. Each participant was assigned a unique study identification number that contained no personally identifiable information. No names or direct identifiers were recorded at any point in data collection or analysis. To protect participant privacy and confidentiality, data were recorded using study IDs only and stored securely in accordance with institutional research guidelines. Only members of the research team had access to the data. Results are presented in aggregate form to prevent identification of individual participants. Participation in this study was voluntary. Participants were not provided with any financial compensation or incentives for participation, and there were no costs associated with participation.

## Results

### Demographics

A total of 40 sports medicine fellowship–trained faculty surgeons were included in the analysis. The mean age of the participants was 43.9 (SD 9.0) years. They had an average of 10.5 (SD 8.3) years in practice and performed approximately 277 (SD 125) arthroscopic cases per year. All were male and 3 (8%) were left-handed. There were no significant differences in demographic data between cohorts from the 2 courses (Table 1).

**Table .** Demographics.

| | Cohort 1 (n=20) | Cohort 2 (n=20) | Overall (n=40) | *P* value[b] |
|---|---|---|---|---|
| Age, mean (SD) | 43.1 (8.38) | 44.7 (9.22) | 43.9 (9.0) | .59 |
| Male gender, n (%) | 20 (100) | 20 (100) | 40 (100) | —[a] |
| Percent fellowship trained, n (%) | 20 (100) | 20 (100) | 40 (100) | — |
| Years in practice, mean (SD) | 9.5 (7.7) | 11.4 (8.5) | 10.5 (8.3) | .48 |
| Cases per year (n), mean (SD) | 275 (130.1) | 279 (117.3) | 277 (125) | .94 |
| Right-handed, n (%) | 18 (90) | 19 (95) | 37 (93) | .55 |

[a]Not applicable.

[b]For age, number of years in practice, and number of cases per year, 2-sided *t* tests were used to test for statistical significance. For handedness, the $\chi^2$ test was used. No differences were statistically significant.

## Probing

For the Probing module, mean completion times were 92.5 (SD 17.6) seconds for the dominant hand and 97.3 (SD 13.6) seconds for the nondominant hand, with no errors recorded and no significant difference between hands (*P*=.35); the benchmark was <95 seconds with no errors.

## Ring Transfer

For the Ring Transfer module, mean completion times were 117.5 (SD 26.7) seconds for the dominant hand and 151.3 (SD 38.1) seconds for the nondominant hand, with mean error rates of 0.6 (SD 0.8) and 1.1 (SD 0.9) dropped rings, respectively; dominant and nondominant hand performance differed significantly (*P*<.001), and the benchmark was <134 seconds with no more than 1 dropped ring.

## Maze

For the Maze module, mean completion times were 89.1 (SD 24.3) seconds for the dominant hand and 108.6 (SD 25.0) seconds for the nondominant hand, with mean error rates of 0.2 (SD 0.4) and 0.1 (SD 0.3) lost balls, respectively; dominant and nondominant hand performance differed significantly (*P*=.02), and the benchmark was <99 seconds with no errors.

## Meniscectomy

For the Meniscectomy module, mean completion times were 73.2 (SD 20.7) seconds for the dominant hand and 63.7 (SD 23.6) seconds for the nondominant hand, with mean error rates of 0.1 (SD 0.3) over- or underresection events for both hands; no significant difference was observed between dominant and nondominant hand performance (*P*=.14), and the benchmark was <68 seconds with no more than 1 area of over- or underresection.

## Suture Passing

For the Suture Passing module, mean completion times were 177.5 (SD 41.2) seconds for the dominant hand and 211.9 (SD 40.9) seconds for the nondominant hand, with mean error distances of 0.5 (SD 0.6) mm and 0.6 (SD 0.8) mm from the target, respectively; dominant and nondominant hand performance differed significantly (*P*<.001), and the benchmark was <195 seconds with no more than 1 mm from the target and no suture anchor unload.

For all modules, benchmark time was defined as the mean performance across dominant and nondominant hands to reflect bilateral task demands. For example, if the right hand took 60 seconds and the left hand took 30 seconds, the benchmark time was 45 seconds. Benchmarks for all modules are summarized in Table 2.

**Table .** Benchmarks.

| Module | Benchmark time (seconds)[a] | Benchmark errors |
|---|---|---|
| Probing module | <95 | No errors |
| Ring Transfer module | <134 | No more than 1 dropped ring |
| Maze module | <99 | No balls off platform |
| Meniscectomy module | <68 | Only 1 area of over- or underresection |
| Suture Passing module | <195 | <1 mm from the target area and no suture anchor unloads |

[a]Benchmark time is the mean performance for both hands. For example, if the right hand takes 60 seconds, and the left hand takes 30 seconds, the time to compare against the benchmark is 45 (90/2) seconds.

## Discussion

### Principal Findings

In this study, we established objective time- and error-based proficiency benchmarks for 5 fundamental arthroscopic skills modules using performance data from fellowship-trained sports medicine faculty members. These benchmarks were derived using a standardized, previously validated simulation-based methodology and were designed to balance efficiency and precision across tasks of increasing technical complexity. Defining proficiency thresholds based on expert performance involves an inherent trade-off. Benchmarks that are too lenient may allow progression without adequate skill consolidation, whereas overly stringent thresholds may discourage practice or be impractical within residency training constraints. By deriving benchmarks from aggregated expert performance and excluding extreme outliers, the present thresholds were designed to balance feasibility with meaningful skill assessment. Continued evaluation of trainee outcomes and future transfer-validity studies will be important to further calibrate these benchmarks.

The FAST program consists of a curriculum, a simulator, and a series of tests, which enable trainees to iteratively practice fundamental skills outside of the operating room and ideally before performing surgery on patients. Training such motor skills outside of the OR is crucial in arthroscopy, which has a steep learning curve [10]. The FAST program's hands-on design is in alignment with the Institute of Medicine's recommendations for graduate medical education, which aim to transition from process-driven training to a more outcome-driven approach. Instead of the assumption that residents acquire their needed skills after completing a certain number of procedures or after a certain amount of time in training, the institute recommends evaluating proficiency in such skills before advancement [11]. The FAST program uses this proficiency-based progression approach to training to encourage "deliberate practice" to improve motor skills [12]. With the establishment of benchmarks for each FAST module, a resident is able to receive real-time objective feedback on their performance and is encouraged to strive to be comparable to experienced arthroscopy surgeons. With the accessibility of the workstations, residents are able to obtain additional practice repetitions and adjust their technique as needed to meet the established standard. Recent evaluations of the FAST workstation have demonstrated measurable differences in performance across postgraduate training levels and institutions, further supporting the need for standardized, objective benchmarks [13].

The FAST program can be integrated as early as the intern year in orthopedic training so that young surgeons can develop foundational psychomotor skills before even touching an arthroscope in the operating room. This early exposure to arthroscopic instruments and techniques can instill confidence in a new resident and their abilities, help them develop good habits early, and potentially decrease patient morbidity. Studies have demonstrated that medical students and residents who train with simulators perform better in the operating room than their counterparts without simulator training, with recent arthroscopy-specific work demonstrating transfer validity to diagnostic knee arthroscopy and meniscectomy performance [14,15]. With improved resident performance in the operating room, a faculty member walking a resident through the case would likely feel more comfortable increasing resident autonomy, and valuable operating room time can be spent teaching and instructing on the unique nuances of a particular case instead of on basic skills.

The FAST workstations are easily accessible, cost-effective, and efficient. They require a simple set-up, and at our institution, they are available in the skills laboratory for all residents to use at any time. Residents can easily independently set up the workstation in the laboratory as long as they have an arthroscopy tower or USB camera with which to practice. Currently, we have each of the testing modules set up in front of a computer monitor with all the instruments available that are necessary to complete the task. We have a QR code that links to a video of the task and instructions mounted on the wall above each of the modules (Figure 1). We use iPevo cameras when a stationary light source is sufficient (Knot Tying) and Sawbones USB cameras when a replica of an arthroscope is required (Ring Transfer). We have found that having easy-to-access instructions and all modules already set up increases use of the simulator and maximizes resident time. We typically introduce FAST during a 2-to-3-hour session with all interns that is proctored by an instructor during their skills month. This familiarizes them with the simulator and modules. They are expected to practice to proficiency and then use the simulators to refine their skills prior to their arthroscopic rotations. The modular design of the FAST program also supports scalability across institutions with varying resource availability and learner profiles, as the workstations can be implemented using differing levels of visualization technology and integrated at multiple stages of training. This flexibility allows programs to adapt FAST implementation to local constraints while maintaining standardized proficiency benchmarks.

**Figure 1.** Simulation laboratory set-up.

Prior papers have aimed to establish clear quantitative benchmarks for simulation training based on the mean performance of an experienced surgeon cohort [16]. One such paper, by Pedowitz et al [17], established the benchmark for the Knot Tying module in the FAST program, which is the last module to be completed in the series. It evaluated the performance of 50 faculty members in attendance at AANA resident arthroscopy courses. The 2 faculty cohorts had an average of 19.3 and 19.9 years of experience in practice. A benchmark was established using data from the more successful of the 2 faculty cohorts and established proficiency as equal to 2 knot failures or less out of 5 knot attempts when using the knot tester workstation. This randomized, prospective study then evaluated a training group of 44 postgraduate year 4 or 5 orthopedic residents. They were divided into 3 subgroups: group A received standard didactic training, group B was allowed additional knot-tying workstation practice, and group C received proficiency-based progression training with the knot-tying workstation. While the aggregate resident knot failure rate of 26% was higher than the 22% knot failure rate of the faculty, resident group C had only an 11% knot failure rate with 94% of residents in this group passing the threshold. This paper both demonstrated a system to establish a benchmark standard and proved that proficiency-based progression training improves the likelihood of meeting said benchmarks. Recent FAST-based educational interventions have demonstrated improvements in objective performance metrics, including arthroscopic knot integrity, when guided, proficiency-based training is used [18]. The arthroscopic knot-tying station is the last in the sequence of the FAST workstation modules as it tests one of the more complex skills in arthroscopy. The purpose of the current investigation was to establish objective benchmarks for the remaining 5 FAST workstation modules: Probing, Ring Transfer, Maze Navigation, Meniscectomy, and Suture Passing.

These benchmarks derived from the performance of fellowship-trained sports medicine faculty members were designed to be both realistic and achievable for residents. The 6 modules progress from basic skills to more complex tasks. The first module, Probing, allows residents to establish basic arthroscopic skills of horizontal control, telescoping, periscoping, and triangulation in order to feel comfortable performing a diagnostic arthroscopy. Gaining familiarity with the arthroscope in this module lays the foundation for establishing visualization in the remainder of the tasks and in the operating room. The second module, Ring Transfer, reinforces these skills and allows the resident to develop proficiency with the arthroscopic grasper, which is used when they are tasked with performing an arthroscopic loose-body removal. The dexterity with the instruments developed in this module is applicable to using arthroscopic instruments to manipulate tissues or implants in all arthroscopic procedures. The next module, Maze Navigation, builds on prior probing skills from the first module and develops tracking skills with nonstationary objects. This skill is necessary to develop before interventional arthroscopy can be performed. While the first 3 modules focus more on skills generalization, the final 3 modules correlate even more closely with specific surgical interventions, thereby focusing more on skills transfer. The fourth module directly simulates completing a partial meniscectomy with a benchmark designed to encourage an adequate amount of resection. During this practice, residents develop more familiarity with a biter and the ability to maneuver more deliberately and precisely. Resection of a paper “meniscus” is a safer alternative to a new resident damaging cartilage with the biter or over-resecting and destabilizing a meniscus tear because they are using a biter for the first time in an actual patient’s knee. The second to last task, Suture Passing, integrates the use of other new instruments, including an antegrade suture passer, a piercing suture passer/retriever, and a suture lasso or suture shuttle. This task’s surgical correlate is an arthroscopic rotator cuff repair or Bankart repair, which are among the more complex arthroscopic procedures. Completion of this task relies on the mastery and coordination of skills obtained in the previous modules. The last module, Knot Tying, allows residents to evaluate the biomechanical integrity of their arthroscopic knots in a setting safer than discovering loose knots in a failed arthroscopic suture repair. In conclusion, a resident’s ability to meet the time constraints and limit their errors for each task

translates into more efficient and accurate performance in the operating room, which promotes patient safety.

In establishing the benchmarks, there was a statistically significant difference in performance between the faculty surgeons' dominant and nondominant hands with the Ring Transfer, Maze, and Suture Passing modules. However, surgeons have to be ambidextrous. Which hand holds the arthroscope and which holds the remaining instruments is often based on the laterality of the procedure. Previous studies have shown that experienced arthroscopic surgeons are more ambidextrous than novices [19]. Specifically, experts demonstrate significantly smaller dominant-nondominant differences in task completion time and error rates on simulated arthroscopic tasks compared with novices, supporting bilateral performance assessment as a marker of skill acquisition. Therefore, the FAST program requires all tasks to be completed with the right hand and the left hand in separate attempts in order to develop skills in both hands regardless of hand dominance. Establishing separate benchmarks for the dominant and the nondominant hand based on the differences seen in the expert cohorts was considered. Aggregated performance metrics are commonly used in established simulation benchmarking frameworks, such as the Objective Structured Assessment of Technical Skill (OSATS) and the Fundamentals of Laparoscopic Surgery program, which define proficiency using composite measures rather than limb-specific performance. This supports the use of bilateral performance means when establishing reproducible and generalizable training benchmarks. That said, the scientific ideal needs to be weighed equally with practicality in this case. The FAST Program puts multiple new demands on trainees and faculty in the programs that adopt it. Ultimately, in order to be successfully adopted, it needs to be easy to implement. Based upon our experience facilitating many FAST courses at the Orthopaedic Learning Center in Chicago and at multiple residency programs, we determined that a single benchmark for each task was necessary.

We also discussed whether to take the mean performance of both hands or to set the benchmark at the level of the dominant hand. We chose the mean performance. This was the best compromise between ensuring adequate ambidexterity while also recognizing that differences exist, even in expert faculty members. The FAST program is designed to enhance surgical skills and promote deliberate, repetitive practice for benchmarks. We were concerned that setting the benchmark as the performance of the dominant hand of the faculty members would have made it significantly harder to achieve for the trainees. The goal was for residents to be able to pass the FAST program by the end of their third year in training, and we decided that the likely significant increase in repetitions required to meet the more stringent benchmark would be frustrating and time consuming for residents who already have many other competing demands. Passing the FAST program at the end of the third year does not mean that residents are proficient at arthroscopy. It means that they have sufficiently practiced and likely have improved their fundamental arthroscopic skills by meeting a minimum threshold of performance. They still have additional years in residency to build upon the skills acquired in the FAST program. Future validation of FAST benchmarks could be pursued through prospective randomized or longitudinal studies comparing proficiency-based FAST training with standard curricula. Studies should be powered to detect differences in operative performance and stratified by postgraduate year to assess differential benefit by training stage. Translation to operating room performance may be measured using objective metrics such as task-specific error rates, time to task completion, and validated global rating scales (eg, OSATS) during comparable arthroscopic procedures. This benchmarking is foundational work that is required to enable such an investigation.

The development of the FAST program benchmarks is one of the factors that allow this to be a self-sufficient progression learning tool, setting it apart from other skills programs. In 2007, the Carolinas Medical Center integrated a laparoscopic skills curriculum into their general surgery residency program [20]. A skills laboratory coordinator needed to be present to record the duration of participation in each task and keep track of errors. In contrast, the FAST program is designed as a web-based program residents can access independently at a time convenient for them. A stopwatch is built into the program to record time to complete a task, and each task has a section to input the number of errors. These factors determine progression to the next module, all of which can be done at the resident's own pace. The paper about the Carolinas Medical Center program by Stefanidis et al [20] describes feedback occurring after each training session or when a resident is called in by the coordinator. With the FAST program, instant feedback is available with the clear benchmarks established for each task.

## Limitations and Future Studies

A limitation of this study was that the included participants were exclusively faculty members teaching at the AANA Fundamentals of Arthroscopic Surgery Residents Course. All participants were male, sports medicine fellowship–trained faculty members, which may limit generalizability across genders, subspecialties, and training environments. Ultimately, this was a convenience sample due to the significant logistical hurdles to obtaining data with another method from a large number of expert orthopedic surgeons. This may have introduced a selection bias for surgeons, and they may not be fully representative of a wider population of sports fellowship–trained surgeons. The included cohort were deemed experts by virtue of their completed training, but there were outliers in performance among the participants. There were 21 (of 320) outlier data points, and these were subsequently excluded from the dataset before formulating the benchmarks.

Another limitation of this study is that it did not demonstrate the feasibility of the FAST program or its transfer validity, necessitating further studies. Further investigation is also needed to determine the average number of repetitions and practice time required to meet the proficiency benchmarks. These results could then be subanalyzed based on training level. This would provide a clearer estimated duration of FAST programs for other orthopedic residency programs interested in integrating the FAST program into their curricula. Further application of the FAST program could also be used to assess whether those who reach proficiency benchmarks perform better than those with

no simulation training on other simulated surgical procedures and in the operating room for comparable surgical procedures. Lastly, further studies are needed to investigate which subset of residents achieve the most benefit from the FAST program as determined by the greatest improvement in arthroscopic skills. This could be used to determine in which year of residency training the FAST program should be integrated.

Benchmarks in this study were derived on a module-specific basis using different expert cohorts and should be interpreted as reference values for individual skills rather than curriculum-level proficiency standards. This design limits direct cross-module comparison but reflects practical constraints of expert data collection and supports initial benchmark development for discrete arthroscopic tasks. Evaluation of learners across the full FAST curriculum will be necessary to assess proficiency progress and program-level validity.

## Conclusions

FAST workstations can be used as proficiency-based learning tools for residents to safely and effectively develop arthroscopic skills outside of the operating room. These benchmarks were established via a method previously validated in surgical simulation and balance precision and efficiency for skills that are considered generalizable and transferable to arthroscopic surgeries.

## Funding

This study was supported by an educational grant from the Arthroscopy Association of North America (AANA) in the amount of $44,000. Study activities were conducted in part at the AANA Orthopaedic Learning Center. The funding source had no role in the design of the study; analysis; interpretation of the data; or in the writing of the manuscript or decision to submit for publication.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Detailed procedural descriptions.
[DOCX File, 2779 KB - periop_v9i1e82723_app1.docx ]

## References

1. Pedowitz RA, Marsh JL. Motor skills training in orthopaedic surgery: a paradigm shift toward a simulation-based educational curriculum. J Am Acad Orthop Surg 2012 Jul;20(7):407-409. [doi: 10.5435/JAAOS-20-07-407] [Medline: 22751159]
2. Derossis AM, Fried GM, Abrahamowicz M, Sigman HH, Barkun JS, Meakins JL. Development of a model for training and evaluation of laparoscopic skills. Am J Surg 1998 Jun;175(6):482-487. [doi: 10.1016/s0002-9610(98)00080-4] [Medline: 9645777]
3. McCluney AL, Vassiliou MC, Kaneva PA, et al. FLS simulator performance predicts intraoperative laparoscopic skill. Surg Endosc 2007 Nov;21(11):1991-1995. [doi: 10.1007/s00464-007-9451-1] [Medline: 17593434]
4. Soper NJ, Fried GM. The fundamentals of laparoscopic surgery: its time has come. Bull Am Coll Surg 2008 Sep;93(9):30-32. [Medline: 19469354]
5. Scott DJ, Bergen PC, Rege RV, et al. Laparoscopic training on bench models: better and more cost effective than operating room experience? J Am Coll Surg 2000 Sep;191(3):272-283. [doi: 10.1016/s1072-7515(00)00339-2] [Medline: 10989902]
6. Seymour NE, Gallagher AG, Roman SA, et al. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. Ann Surg 2002 Oct;236(4):458-463. [doi: 10.1097/00000658-200210000-00008] [Medline: 12368674]
7. Hyltander A, Liljegren E, Rhodin PH, Lönroth H. The transfer of basic skills learned in a laparoscopic simulator to the operating room. Surg Endosc 2002 Sep;16(9):1324-1328. [doi: 10.1007/s00464-001-9184-5] [Medline: 11988802]
8. Korndorffer JR Jr, Dunne JB, Sierra R, Stefanidis D, Touchard CL, Scott DJ. Simulator training for laparoscopic suturing using performance goals translates to the operating room. J Am Coll Surg 2005 Jul;201(1):23-29. [doi: 10.1016/j.jamcollsurg.2005.02.021] [Medline: 15978440]
9. Sroka G, Feldman LS, Vassiliou MC, Kaneva PA, Fayez R, Fried GM. Fundamentals of laparoscopic surgery simulator training to proficiency improves laparoscopic performance in the operating room-a randomized controlled trial. Am J Surg 2010 Jan;199(1):115-120. [doi: 10.1016/j.amjsurg.2009.07.035] [Medline: 20103076]
10. Atesok K, Mabrey JD, Jazrawi LM, Egol KA. Surgical simulation in orthopaedic skills training. J Am Acad Orthop Surg 2012 Jul;20(7):410-422. [doi: 10.5435/JAAOS-20-07-410] [Medline: 22751160]
11. Committee on the Governance and Financing of Graduate Medical Education, Board on Health Care Services, Institute of Medicine. In: Eden J, Berwick DM, Wilensky GR, editors. Graduate Medical Education That Meets the Nation's Health Needs: National Academies Press (US); 2014. [Medline: 25340242]
12. Thomas GW, Johns BD, Marsh JL, Anderson DD. A review of the role of simulation in developing and assessing orthopaedic surgical skills. Iowa Orthop J 2014;34:181-189. [Medline: 25328480]

13. Chapek M, Otlans PT, Buuck T, et al. Resident performance on the Fundamentals of Arthroscopic Surgery Training workstation does not predictably improve with postgraduate year. Arthrosc Sports Med Rehabil 2024 Feb;6(1):100866. [doi: 10.1016/j.asmr.2023.100866] [Medline: 38318395]
14. Palter VN, Grantcharov T, Harvey A, Macrae HM. Ex vivo technical skills training transfers to the operating room and enhances cognitive learning: a randomized controlled trial. Ann Surg 2011 May;253(5):886-889. [doi: 10.1097/SLA.0b013e31821263ec] [Medline: 21394017]
15. Tronchot A, Casy T, Vallee N, et al. Virtual reality simulation training improve diagnostic knee arthroscopy and meniscectomy skills: a prospective transfer validity study. J Exp Orthop 2023 Dec 14;10(1):138. [doi: 10.1186/s40634-023-00688-8] [Medline: 38095746]
16. Gallagher AG, O'Sullivan GC. Fundamentals of Surgical Simulation: Principles and Practice: Springer; 2011.
17. Pedowitz RA, Nicandri GT, Angelo RL, Ryu RKN, Gallagher AG. Objective assessment of knot-tying proficiency with the Fundamentals of Arthroscopic Surgery Training program workstation and knot tester. Arthroscopy 2015 Oct;31(10):1872-1879. [doi: 10.1016/j.arthro.2015.06.021] [Medline: 26298642]
18. Farooq H, Gaetano A, Salazar D, Garbis N. Guided education and the Fundamentals of Arthroscopy Surgery Training (FAST) workstation improve surgical resident knot-tying skills. Arthrosc Sports Med Rehabil 2025 Jun;7(3):101110. [doi: 10.1016/j.asmr.2025.101110] [Medline: 40692921]
19. Pedowitz R, Nicandri G, Tuchschmid S. Asymmetry in dominant / non-dominant hand performance differentiates novices from experts on an arthroscopy virtual reality serious game. Stud Health Technol Inform 2016;220:289-294. [Medline: 27046594]
20. Stefanidis D, Acker CE, Swiderski D, Heniford BT, Greene FL. Challenges during the implementation of a laparoscopic skills curriculum in a busy general surgery residency program. J Surg Educ 2008;65(1):4-7. [doi: 10.1016/j.jsurg.2007.11.009] [Medline: 18308276]

## Abbreviations

**AANA:** Arthroscopy Association of North America
**FAST:** Fundamentals of Arthroscopic Surgery Training
**OSATS:** Objective Structured Assessment of Technical Skill

*Edited by N Rohatgi; submitted 25.Aug.2025; peer-reviewed by D Demirel, W Li; revised version received 22.Feb.2026; accepted 25.Feb.2026; published 08.Apr.2026.*

*Please cite as:*
*Davis E, Caraet B, Pedowitz R, Nicandri G*
*Benchmark Development for Fundamental Arthroscopic Skills Using a Simulation-Based Training Program: Observational Study*
*JMIR Perioper Med 2026;9:e82723*
*URL: https://periop.jmir.org/2026/1/e82723*
*doi:10.2196/82723*

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org