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 8 (2025)   ISSN 2561-9128   Editor in Chief: Nidhi Rohatgi, MS, MD, SFHM

Contents

Original Papers

Original Paper

Reducing Greenhouse Gas Emissions and Modifying Nitrous Oxide Delivery at Stanford: Observational, Pilot Intervention Study

Eric P Kraybill[1], BS; David Chen[1], MD; Saadat Khan[1], MEng; Praveen Kalra[1], MBBS, MD

Stanford Hospital, Stanford, CA, United States

Corresponding Author:
Praveen Kalra, MBBS, MD
Stanford Hospital
300 Pasteur Drive
Suite H3580
Stanford, CA, 94305
United States
Phone: 1 650 725 6412
Email: pkalra@stanford.edu

Abstract

Background: Inhalational anesthetic agents are a major source of potent greenhouse gases in the medical sector, and reducing their emissions is a readily addressable goal. Nitrous oxide (N_2O) has a long environmental half-life relative to carbon dioxide combined with a low clinical potency, leading to relatively large amounts of N_2O being stored in cryogenic tanks and H cylinders for use, increasing the chance of pollution through leaks. Building on previous findings, Stanford Health Care's (SHC's) N_2O emissions were analyzed at 2 campuses and targeted for waste reduction as a precursor to system-wide reductions.

Objective: We aimed to determine the extent of N_2O pollution at SHC and subsequently whether using E-cylinders for N_2O storage and delivery at the point of care in SHC's ambulatory surgery centers could reduce system-wide emissions.

Methods: In phase 1, total SHC (Palo Alto, California) N_2O purchase data for calendar year 2022 were collected and compared (volume and cost) to total Palo Alto clinical delivery data using Epic electronic health records. In phase 2, a pilot study was conducted in the 8 operating rooms of SHC campus A (Redwood City). The central N_2O pipelines were disconnected, and E-cylinders were used in each operating room. E-cylinders were weighed before and after use on a weekly basis for comparison to Epic N_2O delivery data over a 5-week period. In phase 3, after successful implementation, the same methodology was applied to campus B, one of 3 facilities in Palo Alto.

Results: In phase 1, total N_2O purchased in 2022 was 8,217,449 L (33,201.8 lbs) at a total cost of US $63,298. Of this, only 780,882.2 L (9.5%) of N_2O was delivered to patients, with 7,436,566.8 L (90.5%) or US $57,285 worth lost or wasted. In phase 2, the total mass of N_2O use from E-cylinders was 7.4 lbs (1 lb N_2O=247.5 L) or 1831.5 L at campus A. Epic data showed that the total N_2O volume delivered was 1839.3 L (7.4 lbs). In phase 3, the total mass of N_2O use from E-cylinders was 10.4 lbs or 2574 L at campus B (confirming reliability within error propagation margins). Epic data showed that the total N_2O volume delivered was 2840.3 L (11.5 lbs). Over phases 2 and 3, total use for campuses A and B was less than the volume of 3 E-cylinders (1 E-cylinder=1590 L).

Conclusions: Converting N_2O delivery from centralized storage to point-of-care E-cylinders dramatically reduced waste and expense with no detriment to patient care. Our results provide strong evidence for analyzing N_2O storage in health care systems that rely on centralized storage, and consideration of E-cylinder implementation to reduce emissions. The reduction in N_2O waste will help meet SHC's goal of reducing scope 1 and 2 emissions by 50% before 2030.

(JMIR Perioper Med 2025;8:e64921) doi:10.2196/64921

KEYWORDS

anesthetic gases; emissions; green house gas; sustainability; pilot study; electronic health record; implementation; nitrous oxide; global warming

Introduction

Reducing greenhouse gas (GHG) emissions is a priority that must be addressed to reduce climate change and its negative impacts on earth and its inhabitants. The US Environmental Protection Agency (EPA) classifies GHG emissions into different categories, with scope 1 emissions defined as direct GHG emissions from sources that are controlled by organizations, including health care systems, and scope 2 emissions being indirect GHG emissions associated with the purchase of electricity, heat, steam, or cooling [1]. Stanford Health Care (SHC) has signed on to the US Department of Health and Human Services' pledge to reduce its scope 1 and 2 emissions by 50% by 2030 [2]. Within the medical sector, inhalational anesthetic gases that are directly released into the atmosphere are a major source of potent GHGs. Thus, there is a fertile opportunity to reduce GHGs by reducing the emission of anesthetic gases [3]. By collecting annual emissions data within the SHC system, improvements to sustainability and infrastructure could be explored.

Global warming potential (GWP) represents the energy a gas is able to absorb relative to carbon dioxide (CO_2), with a larger GWP representing increased planetary warming [4]. The environmental impacts of 2 inhaled anesthetic gases over a 100-year period (ie, global warming potential of GHGs over a 100-year period [GWP100]) are particularly relevant: desflurane, a volatile halogenated agent with particularly high GWP100 of 2540, and nitrous oxide (N_2O) with a lower GWP100 of 298 but used in much higher volumes than other anesthetic gases, and with longer half-life compared to CO_2, leading to lasting environmental consequences [5]. Further, because of its low clinical potency, large amounts of N_2O must be stored for use, increasing the chance of pollution through leaks. Centrally piped cryogenic liquid, centrally piped gas, and portable E-cylinders are the standard options for delivering N_2O [6]. Miles of pipes and innumerable valves in centrally piped systems lead to an abundance of leaks, contributing to excessive loss and waste [6]. While desflurane has already been discontinued from routine clinical use at SHC, we aimed to determine the degree to which N_2O emissions could be reduced and waste prevented, building on prior studies highlighting the waste of N_2O in other institutions [7].

Methods

Phase 1

To begin investigating N_2O emissions, purchase data (volume and cost) were collected and compared to total use data (clinical delivery) using the Epic SlicerDicer tool, part of the Epic electronic health record (EHR) [8]. Epic yearly clinical use data for N_2O are available per clinical service in the SHC's operating rooms. Gas losses in the system can be estimated by comparing documented gas delivery at the point of care with the volume of N_2O purchased. Initial data analysis revealed a drastic amount of lost N_2O, leading us to perform a pilot study (phase 2, E-cylinder implementation) to enable remediation aimed at reducing N_2O emissions.

Phase 2

Using the Institute for Healthcare Improvement framework of "Plan, Do, Study, Act" for performance improvement [9], a pilot study was conducted in the 8 operating rooms of the SHC campus in Redwood City, California (campus A). E-cylinder canisters were deployed in each operating room and all central N_2O pipelines were disconnected. EHR documentation of gas delivered in liters (volume) was compared to measured E-cylinder mass. To verify use and track N_2O leaving each tank, the E-cylinders were weighed before and after use on a weekly basis with the difference in mass converted to volume (liters). Since the measured pressure remains the same as long as liquid remains in the cylinders, pressure differences cannot be used for measuring N_2O flow until only gas is left (at which point the pressure drop correlates with the amount of gas being removed) [10]. By using the conversion of 1 lb (0.45 kg) of N_2O being equal to 247.5 L [6], the volume of N_2O dispensed could be calculated. Total calculated volume leaving the E-cylinders based on measured mass was compared to total volume delivered according to Epic data.

Phase 3

Following the results of phase 2, a secondary study was conducted in 16 operating rooms at Blake Wilbur Drive Palo Alto, California (campus B). Phase 3 used the same methodology as phase 2 over a 3-week period.

Ethical Considerations

Due to the nature of the research and institutional approval, no IRB approval was necessary. No identifying patient data was used as we only measured nitrous oxide gas delivery and utilization.

Results

Phase 1

According to the Stanford Medicine Sustainability Program Office [2], the annual Palo Alto SHC 2022 Scope 1 emissions were 19,374 $MTCO_2e$ (metric ton of CO_2 equivalent, the standard unit for comparing different GHGs to quantify their environmental impact and GWP) of which medical gases (including N_2O, CO_2, sevoflurane, and isoflurane) represented 4862 $MTCO_2e$. N_2O contributed 4590 $MTCO_2e$ of the medical gases. Thus, medical gases account for 25.1% of all SHC scope 1 emissions, and N_2O alone accounts for 94.4% of those emissions (or 23.7% of the total).

Annual clinical usage of N_2O in 2022 per Epic data (Table 1) was 780,882.2 L (3155.1 lbs or 1431.1 kg), with the greatest use being for orthopedic surgery, general surgery, and neurosurgery cases. However, the total amount of N_2O purchased was 8,217,449 L (33,201.8 lbs or 15,060.1 kg), at a total cost of US $63,298. Thus, only 9.5% of the total purchased N_2O was actually delivered to patients, and 90.5% (or US $57,285 worth) was wasted.

Table 1. Annualized data comparing centralized N_2O system to hypothetical E-cylinders for Stanford Health Care (SHC; all campuses).

	Amount purchased (L)	Cost (US $)	Amount used (L)	Amount lost as waste (L)
Centralized system	8,217,449	63,298	780,882.2	7,436,566.8
E-cylinders	780,882.2[a]	6015	780,882.2	0[b]

[a]Amount needed to purchase with zero surplus based on use data under experimental conditions.

[b]Annualized E-cylinder data are extrapolated from experimental conditions; real-world conditions may vary.

With these data indicating a loss of greater than 90% between storage tanks and clinical use, a highly inefficient storage and pipeline system was recognized. The proposed solution (for phase 2 of the study) was to decommission the storage tanks and pipelines and switch to portable E-cylinders that stored and delivered N_2O at the point of care.

Phase 2

The change in mass of the E-cylinders indicated that N_2O use at campus A totaled 7.4 lbs (3.4 kg), or a volume of 1831.5 L, over the 5-week study period. Epic data showed total N_2O volume delivered to be 1839.3 L calculated to 7.4 lbs (3.4 kg; consistent with the measured 7.4 lbs). Using the standard of 1 E-cylinder=1590 L or 6.4 lbs (2.9 kg) [11], total use equaled 1.16 E-cylinders.

Phase 3

The E-cylinder change in mass indicated that N_2O use at campus B totaled 10.4 lbs (4.7 kg), or 2574 L, over the 3-week data collection period. Epic data showed total N_2O volume delivered to be 2840.3 L calculated to 11.5 lbs (5.2 kg; compared to the measured 10.4 lbs, which would be equivalent to 1.63 E-cylinders) [11].

Discussion

Principal Findings

Results from phase 1 corroborate findings from previous studies in the United Kingdom and Portland, Oregon [12,13], which reveal excessive waste from centralized storage of N_2O and pipe systems for delivery. Phases 2 and 3 of this study, from 2 different SHC campuses, demonstrate the efficient, cost-effective elimination of waste through substitution of E-cylinders with storage and delivery at the point of care. In phases 2 and 3, avoidable N_2O emissions were almost completely eliminated (Multimedia Appendix 1). The discrepancy between actual weighed N_2O and Epic-reported use for campus A was 7.8 L, or <0.1 lb (<0.1 kg). Campus B had a greater discrepancy with the difference in actual weighed N_2O and Epic-reported use being 266.3 L, or 1.1 lbs (0.45 kg). The amount of gas delivered according to the EHR was greater than the amount actually measured at the source, potentially accounted for by limited precision of the scales used to weigh the E-cylinders (only to 0.1-lb increments), or accidental reconnection of N_2O pipelines in one operating room during phase 3. This issue was detected after 1 week and immediately rectified.

E-cylinders provide an efficient and effective solution, but they hold limitations. E-cylinders must be stored properly to ensure that they do not present a catalyst in the event of a fire [14]. However, no policy implementation is required as E-cylinders are already in use in operating rooms and costs associated with storage can be offset by the N_2O saved. Ready accessibility, lower cost, reduced supply chain issues, and efficiency of E-cylinders far outweigh the abovementioned disadvantages.

Limitations

The limitations of this study include the fact that real-world use and waste may vary from our experimental conditions, likely incurring greater losses. If e-cylinder valves are accidentally left open, losses may simulate those from centralized pipelines until the valve is closed [6] or the E-cylinder is emptied. The amount of N_2O to be purchased would need to be greater than the amount used in our example (Table 1), to provide surplus in the E-cylinders as well as spare E-cylinders. Prospective estimates of volume when making a purchase order would likely exceed actual use. Both recording and documentation of N_2O readings and the scale measurements are susceptible to error.

Conclusions

Converting the delivery of N_2O from centralized storage to point-of-care E-cylinders has dramatically reduced waste and expense with no detriment to patient care. Stanford's pledge to reduce scope 1 and 2 emissions by 50% can be achieved and even surpassed if this practice is changed in all SHC locations. The introduction of E-cylinders will provide a nondisruptive means for immediately decreasing emissions while continuing to provide optimal anesthetic care. Pilot studies throughout Stanford's campuses continue, with the goal of removing the centralized N_2O system and switching to E-cylinders at other sites, thereby significantly reducing anesthetic GHG emissions. Efforts to reduce GHG emissions may begin locally but have applications globally. Reducing the anesthetic carbon footprint of health care organizations is necessary for our planet and can begin with the reduction of wasteful emissions.

Acknowledgments

We would like to acknowledge the Stanford Sustainability Planning Office for their support throughout the project.

Authors' Contributions

EPK conducted the data analysis and drafted and edited the manuscript. DC conducted the data analysis and edited the manuscript. SK collected and analyzed the data. PK conceptualized the study, conducted and analyzed the data, and edited the manuscript. PK and SK (SaadatKhan@stanfordhealthcare.org) are co-corresponding authors.

Conflicts of Interest

PK is an associate editor for *JMIR Perioperative Medicine*.

Multimedia Appendix 1

Reduction in N_2O emissions per metric ton of CO_2 equivalents by switching from the original central supply to portable supply E-cylinder storage.

[PNG File , 58 KB - periop_v8i1e64921_app1.png]

References

1. Scope 1 and Scope 2 Inventory Guidance. United States Environmental Protection Agency. 2020. URL: https://www.epa.gov/climateleadership/scope-1-and-scope-2-inventory-guidance [accessed 2024-12-16]
2. Stanford Medicine. Our Sustainability Commitment. Stanford Health Care. URL: https://stanfordhealthcare.org/sustainability-program-office/sustainability-program-office/what-we-do/our-sustainability-commitment.html [accessed 2024-12-16]
3. Chesebro BB, Gandhi S. Mitigating the systemic loss of nitrous oxide: a narrative review and data-driven practice analysis. Br J Anaesth 2024 Dec;133(6):1413-1418. [doi: 10.1016/j.bja.2024.08.028] [Medline: 39322471]
4. Understanding Global Warming Potentials. United States Environmental Protection Agency. 2016. URL: https://www.epa.gov/ghgemissions/understanding-global-warming-potentials [accessed 2024-12-16]
5. Sulbaek Andersen MP, Nielsen OJ, Wallington TJ, Karpichev B, Sander SP. Medical intelligence article: assessing the impact on global climate from general anesthetic gases. Anesth Analg 2012 May;114(5):1081-1085. [doi: 10.1213/ANE.0b013e31824d6150] [Medline: 22492189]
6. Collaborating to prevent nitrous oxide waste in medical gas systems. Practice Greenhealth. URL: https://practicegreenhealth.org/tools-and-resources/collaborating-prevent-nitrous-oxide-waste-medical-gas-systems [accessed 2024-12-16]
7. Seglenieks R, Wong A, Pearson F, McGain F. Discrepancy between procurement and clinical use of nitrous oxide: waste not, want not. Br J Anaesth 2022 Jan;128(1):e32-e34 [FREE Full text] [doi: 10.1016/j.bja.2021.10.021] [Medline: 34802695]
8. Our Software. Epic Systems Corporation. URL: https://www.epic.com/software/ [accessed 2024-12-16]
9. How to Improve: Model for Improvement. Institute for Healthcare Improvement. URL: https://www.ihi.org/resources/how-to-improve [accessed 2024-12-16]
10. Medical Gases: Storage and Supply. Anesthesia Key. 2019. URL: https://aneskey.com/medical-gases-storage-and-supply-2/ [accessed 2024-12-16]
11. Rose G, McLarney J. Pneumatic Systems. In: Anesthesia Equipment Simplified. New York, NY: McGraw-Hill Education; 2014.
12. Devlin-Hegedus JA, McGain F, Harris RD, Sherman JD. Action guidance for addressing pollution from inhalational anaesthetics. Anaesthesia 2022 Sep 21;77(9):1023-1029 [FREE Full text] [doi: 10.1111/anae.15785] [Medline: 35729804]
13. Sherman J. It's time hospitals abandon nitrous oxide pipes. ASA Monitor 2024;88:33 [FREE Full text] [doi: 10.1097/01.ASM.0001006828.24359.cd]
14. Nitrous oxide. CAMEO Chemicals. URL: https://cameochemicals.noaa.gov/chemical/8909#:~:text=It%20is%20noncombustible%20but%20it,to%20rupture%20violently%20and%20rocket [accessed 2024-12-11]

Abbreviations

EHR: electronic health record
EPA: Environmental Protection Agency
GHG: greenhouse gas
GWP: global warming potential
GWP100: global warming potential of GHGs over a 100-year period
$MTCO_2e$: metric ton of carbon dioxide equivalent
N_2O: nitrous oxide
SHC: Stanford Health Care

Edited by T Aslanidis; submitted 09.08.24; peer-reviewed by A Maleki, LM Western; comments to author 07.10.24; revised version received 27.11.24; accepted 02.12.24; published 09.01.25.

Please cite as:
Kraybill EP, Chen D, Khan S, Kalra P
Reducing Greenhouse Gas Emissions and Modifying Nitrous Oxide Delivery at Stanford: Observational, Pilot Intervention Study
JMIR Perioper Med 2025;8:e64921
URL: https://periop.jmir.org/2025/1/e64921
doi:10.2196/64921
PMID:

Original Paper

A Patient-Oriented Implementation Strategy for a Perioperative mHealth Intervention: Feasibility Cohort Study

Daan Toben[1,2], MSc; Astrid de Wind[1,2], PhD; Eva van der Meij[2,3], PhD; Judith A F Huirne[2,3,4], PhD; Johannes R Anema[2], PhD

[1]Department of Public and Occupational Health, Amsterdam UMC location Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

[2]Societal Participation & Health, Amsterdam Public Health, Amsterdam, The Netherlands

[3]Department of Obstetrics and Gynaecology, Amsterdam UMC location University of Amsterdam, Amsterdam, The Netherlands

[4]Department of Public Health, Amsterdam University Medical Center, University of Amsterdam, Amsterdam, The Netherlands

Corresponding Author:
Daan Toben, MSc
Department of Public and Occupational Health
Amsterdam UMC location Vrije Universiteit Amsterdam
Van der Boechorststraat 7
Amsterdam, 1081 BT
The Netherlands
Phone: 31 647861694
Email: d.j.toben@amsterdamumc.nl

Abstract

Background: Day surgery is being increasingly implemented across Europe, driven in part by capacity problems. Patients recovering at home could benefit from tools tailored to their new care setting to effectively manage their convalescence. The mHealth application ikHerstel is one such tool, but although it administers its functions in the home, its implementation hinges on health care professionals within the hospital.

Objective: We conducted a feasibility study of an additional patient-oriented implementation strategy for ikHerstel. This strategy aimed to empower patients to access and use ikHerstel independently, in contrast to implementation as usual, which hinges on the health care professional acting as gatekeeper. Our research question was "How well are patients able to use ikHerstel independently of their health care professional?"

Methods: We investigated the implementation strategy in terms of its recruitment, reach, dose delivered, dose received, and fidelity. Patients with a recent or prospective elective surgery were recruited using a wide array of materials to simulate patient-oriented dissemination of ikHerstel. Data were collected through web-based surveys. Descriptive analysis and open coding were used to analyze the data.

Results: Recruitment yielded 213 registrations, with 55 patients ultimately included in the study. The sample was characterized by patients undergoing abdominal surgery, with high literacy and above average digital health literacy, and included an overrepresentation of women (48/55, 87%). The implementation strategy had a reach of 81% (63/78), with 87% (55/67) of patients creating a recovery plan. Patients were satisfied with their independent use of ikHerstel, rating it an average 7.0 (SD 1.9) of 10, and 54% (29/54) of patients explicitly reported no difficulties in using it. A major concern of the implementation strategy was conflicts in recommendations between ikHerstel and the health care professionals, as well as the resulting feelings of insecurity experienced by patients.

Conclusions: In this small feasibility study, most patients were satisfied with the patient-oriented implementation strategy. However, the lack of involvement of health care professionals due to the strategy contributed to patient concerns regarding conflicting recommendations between ikHerstel and health care professionals.

(JMIR Perioper Med 2025;8:e58878) doi:10.2196/58878

KEYWORDS

perioperative care; recovery; feasibility; convalescence; patient-oriented; surgery; perioperative; eHealth; mHealth; tailor; customize; patient care; digital intervention; health intervention; patient education; surgical care; hospital care; digital health; perioperative medicine; elective surgery; technology; caregiver; mobile app; digital care

Introduction

Day surgery—defined as admittance to and discharge from a hospital within 24 hours following surgery—has seen a marked increase in Organisation for Economic Co-operation and Development member countries over the past decades [1]. The appeal of day surgery derives from multiple factors, including its reduced cost, decreased morbidity and mortality, and high levels of patient satisfaction [2-6]. When it comes to postsurgical recovery, however, the reports are more nuanced. Tran et al [7] showed how 1 in 3 patients exhibit suboptimal recovery trajectories following day surgery. Patients recovering at home describe feelings of insecurity, an experience moderated by the timely provision of information, professional support, and expectation management [4,8-12]. mHealth interventions have been shown to be effective when it comes to targeting these domains and their use in the perioperative setting is well appreciated by patients [13,14]. In the Netherlands specifically, the Patient Journey app has been shown to improve postoperative outcomes for patients with musculoskeletal disorders [15].

Similarly, the mHealth intervention ikHerstel (meaning "I recover" in Dutch) is a tool designed to support patients undergoing abdominal surgery during their perioperative period. The intervention's ability to speed up postoperative recovery, reduce pain, and improve patients' quality of life has been established in previous studies [12,16-18]. However, its implementation occurs on the level of the hospital ward, and it hinges on the involvement of health care professionals within the ward, who act as both distributors of the intervention and instructors of patients. This strategy features benefits as well as challenges: health care professionals are well situated to select eligible patients and can improve adherence to treatment when they use effective communication strategies [19,20]. However, at the time of publishing, the intervention has been implemented in only 10% of hospitals in the Netherlands. Wider implementation is hampered by, among other factors, financial barriers present in the Dutch health care system that make upscaling of telemonitoring interventions in general a difficult enterprise [21]. This limits ikHerstel's reach, leaving patients bereft of its aforementioned benefits.

In this feasibility study, we explored a patient-oriented implementation strategy for ikHerstel that aimed to circumvent this hospital-level barrier by targeting patients directly. If successful, this strategy could operate in addition to implementation as usual, with reimbursement flowing from health insurers to patients. We therefore aimed to evaluate whether it would be successful in increasing the intervention's reach and whether patients, once reached, were able to use ikHerstel independently from their health care professional.

Methods

Ethical Considerations

Approval for the study was granted by the medical ethics committee of Amsterdam University Medical Center on May 31, 2022 (2022.0224). Informed consent was obtained through postal mail and patients were informed of their ability to opt out of participation in the study at any time. Patients were provided with access to ikHerstel free of charge but were not offered any remuneration for their participation in the study. Data were deidentified by the coordinating researcher, and patients were labeled using random strings. The patient identification keys were kept in a separate location from the data.

Study Setting

We conducted a prospective study assessing the feasibility of a patient-oriented implementation strategy for the ikHerstel mHealth intervention. Our assessment was performed based on the model of Steckler and Linnan [22]; its outcomes were reach, dose delivered, dose received, fidelity, and recruitment. In consultation with health insurers and a patient interest group, we aimed to include 100 perioperative patients representing the theoretical user base of the ikHerstel app, that is, any patients who were theoretically able to access the app and use it in such a way as to manage their own recovery, regardless of age, gender, nationality, literacy, digital literacy, or health literacy. Recruitment started in September 2022 and lasted through September 2023.

Inclusion and Exclusion Criteria

Patients were eligible for inclusion if they were older than 18 years, proficient in the Dutch language, and prospective recipients of one of the following elective surgical procedures: laparoscopic or abdominal hysterectomy, laparoscopic cholecystectomy, open or laparoscopic inguinal hernia surgery, or laparoscopic adnexal surgery. Patients were excluded if the date of their surgery was ≥14 days prior to inclusion, they were undergoing a combination of surgeries, they had comorbidities that invalidated the convalescence recommendations provided by ikHerstel, they were undergoing oncological surgery, or they were receiving care from a hospital that had already implemented ikHerstel.

Intervention and Procedure

ikHerstel was developed in collaboration with health care professionals of a diverse background. Its development process has been described previously [23]. An overview of the current functions and layout of ikHerstel is provided in Multimedia Appendix 1. Its aim is to prepare patients and manage their expectations preoperatively and to support them in recovery of the daily functions of life postoperatively [23]. Each patient received the ikHerstel intervention in addition to usual care. Patients were able to interact with the intervention in the form of a mobile app, which they used up to the point of their total recovery. They were provided with personal accounts in which they constructed their recovery plan through goal attainment by selecting 8 personal activities from a list of 31 to constitute their most important recovery goals. In this way, one patient might create a plan focused on performing tasks around the house while another might create one centered on regaining the ability to run long distances. Patients monitored their recovery plan through the mobile app: they were asked to indicate when they were able to perform each of the activities in their plan. The total postoperative recovery was visible as a percentage within the app. Additionally, educational material about recovery was provided to patients in the form of text and video animations through the app's library screen.

Implementation Strategy as Usual

In its current form, ikHerstel's implementation strategy hinges on health care professionals, who recruit eligible patients, introduce them to the app and its potential benefits, and provide them with access by creating a personal account. This final step is particularly crucial, as patients cannot access ikHerstel without an account, and health care professionals preload each account with recovery-related data specific to the patient's surgical procedure. Implementation occurs at the level of the hospital ward. A medical liaison associated with ikHerstel trains the ward's staff in the app's use and goals and in carrying out support tasks like creating patient accounts. The hospital ward is also provided with a web portal that mediates these administrative functions, allows for monitoring of each patient's recovery, and provides health care professionals with organizational support.

Patient-Oriented Implementation Strategy

The patient-oriented implementation strategy piloted in this study circumvented health care professionals, relying instead on patients to sign up and use ikHerstel independently. Health care professionals did not have access to the app or the web portal. Instead, these responsibilities were assigned to the coordinating researcher as a placeholder for the support staff of the ikHerstel spinoff company. During the course of the study, the coordinating researcher created patients' accounts and loaded them with surgery-related data based on information provided by the patients. Patient monitoring through the web portal was not performed. In case of questions concerning ikHerstel, patients were directed to the coordinating researcher, whose contact details were provided. Patients with medical questions were directed by the researcher to consult their health care professional. This highlights the key role still reserved for health care professionals in this patient-oriented implementation, as they retained responsibility for care of their patients, including monitoring for adverse outcomes. Accordingly, patients were informed that their health care professional held final authority over the content and provision of care. Figure 1 illustrates the differences between the implementation strategies. Table 1 presents an overview of the recruitment tools that were used, distinguishing between hospital-independent and -dependent tools.

With the exception of the magazine advertisements, all advertisements followed the same basic design, created with low-literacy patients in mind. An example is provided in Multimedia Appendix 2. These materials were distributed to patients in hospitals, on patient fora, on webpages of patient interest groups, in patient magazines, through internet search engine advertisements, and within patient groups on social media. Each advertisement linked to a web portal where patients were informed of the study and asked to leave their contact details. Patients were subsequently contacted via telephone by the coordinating researcher, who provided further information and performed screening on the basis of the inclusion and exclusion criteria.

Figure 1. Schematic representation of the differences between implementation as usual and the patient-oriented implementation.

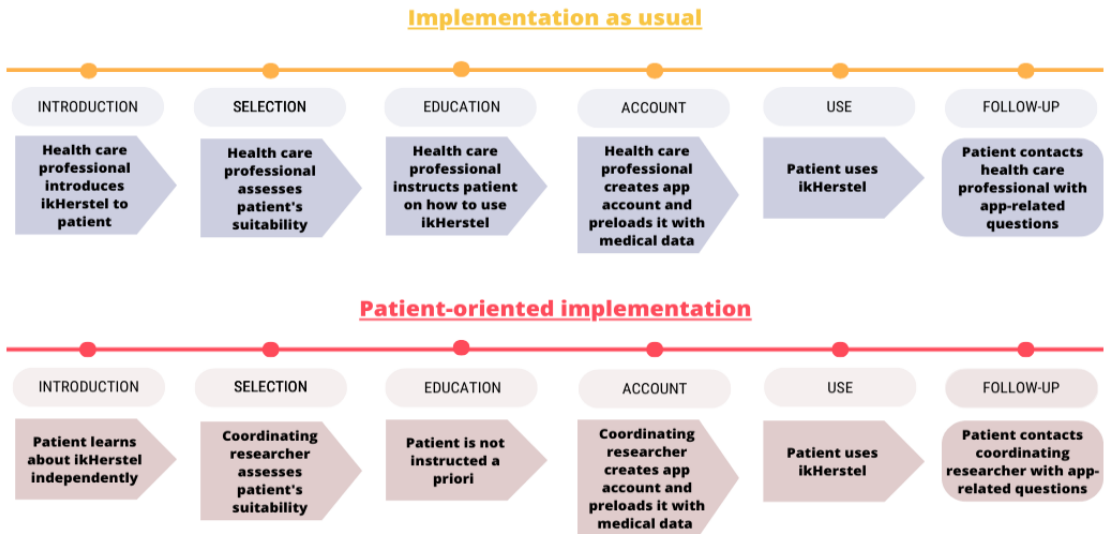


Table 1. Materials used for study recruitment and the frequency of their use, split into hospital-dependent and -independent tools.

Materials	Frequency of use, n
Hospital-independent	
Forum advertisements	15
Webpage advertisements	2
Internet search engine advertisements	1
Social media advertisements	4
Magazine advertisements	2
Hospital-dependent	
Flyers	11
Posters	10
Business cards	6
Electronic displays	5
Hospital staff	2

Data Collection

Data were collected through a set of 4 digital surveys constructed, distributed, and maintained through Survalyzer (Survalyzer AG). A baseline survey (T_0) was used to collect demographic data. Follow-up surveys were distributed to patients at T_1 (3 weeks), T_2 (6 weeks), and T_3 (12 weeks) after surgery to collect data on the user experience.

Background Factors and Implementation Outcomes

Demographic data included socioeconomic factors like age, sex, and education level, which is aligned with a previous study by van der Meij et al [24]. Demographics also included a measure of patients' traditional literacy, operationalized on the basis of the Diagnostic Illiteracy Scale, where a score of 14 points or higher constitutes a risk of the individual being illiterate [25]. Digital literacy was operationalized using patient self-assessment and a scanning tool (Quickscan) developed for physicians by the Dutch patient advocate organization Pharos, which characterizes patients as digitally unskilled with a score of 10 points or higher [26].

The model by Steckler and Linnan [22], commonly used in public health, describes the evaluation of implementation outcomes as a concatenated appraisal of an intervention's context, reach, dose delivered, dose received, fidelity, and recruitment. Operationalization of these outcomes was performed similarly to previous process evaluations of ikHerstel to facilitate comparison [24,27]. We omitted the aspect of fidelity, as the app does not deviate from protocol in its delivery of the intervention. We also omitted context, as this is described in earlier publications, as well as the aspect of implementation, as we judged its transformation of the other aspects into a summative score to be a bad fit for our study. We also evaluated the recruitment tools and their channels (hospital dependent vs independent) in terms of their effectiveness in recruiting eligible patients to use the app. To compute this count, we asked patients to state how they were informed about the study.

We measured patient attitudes in alignment with the patient-oriented character of the implementation strategy and for comparison with previous research [24,27]. We operationalized patient attitudes as patients' self-reported satisfaction rating and their experienced barriers to use. We additionally measured patient attitudes using the unified theory of acceptance and use of technology 2 (UTAUT2), developed by Venkatesh et al [28]. Briefly, this framework describes an individual's intention to use a technology as being determined by 7 constructs: performance expectancy, effort expectancy, social influence, facilitating conditions, hedonic motivation, price value, and habit. Social influence and hedonic motivation were deemed less relevant to ikHerstel's context and thus were not included. Relevant UTAUT2 survey items were selected by the researchers, adapted to the research context, and translated into Dutch. Response categories followed a 4-point Likert scale centered on agreement. The resultant survey is provided in Multimedia Appendix 3. A full overview of the study's outcomes and their operationalization is presented in Table 2.

Table 2. Operationalization of implementation outcomes and patient attitudes.

	Description	Operationalization
Implementation outcomes[a]		
Reach	The proportion of the intended target audience that participated in the study	Numerator: number of patients who met the inclusion criteria and signed an informed consent form; denominator: number of patients who met the inclusion criteria, regardless of their eventual participation in the study
Dose delivered	The number or amount of intended units of the intervention provided to the study population	Numerator: number of patients who were provided with an account for the ikHerstel app; denominator: number of patients who met the inclusion criteria and signed an informed consent form
Dose received	The extent to which participants actively engaged with, interacted with, were receptive to, or used the intervention	Numerator: number of patients who activated their ikHerstel account, created a recovery plan, and used the app on a weekly basis; denominator: number of patients who were provided with an account for the ikHerstel app
Recruitment	The effectiveness of the procedures used to attract participants	An appraisal of the effectiveness of each recruitment medium (hospital dependent vs independent) and tool in terms of the number of inclusions versus registrations they produced
Patient attitudes		
Patient satisfaction	—[b]	Patient satisfaction, assessed through a self-reported score between 0 and 10
Barriers to use	—	Five open questions: • What did you like about using ikHerstel? • What makes using ikHerstel easy? • What did you dislike about using ikHerstel? • What makes using ikHerstel difficult? • Do you have any other comments about the ikHerstel app?
Performance expectancy[c]	The degree to which using the technology will provide benefits to consumers	The degree to which patients view ikHerstel as being able to beneficially affect their postsurgical recovery; operationalized as 3 self-reported items, scored using a 1-4 Likert scale
Effort expectancy[c]	The degree of ease associated with consumers' use of the technology	The degree to which patients feel using ikHerstel is simple and straightforward; operationalized as 3 self-reported items, scored using a 1-4 Likert scale
Facilitating conditions[c]	Consumers' perceptions of the resources and support available to perform a behavior	The degree to which patients feel they are supported in their use of ikHerstel; operationalized as 2 self-reported items, scored using a 1-4 Likert scale
Price value[c]	Consumers' cognitive tradeoff between the perceived benefits of the technology and the monetary cost for using it	The degree to which patients are willing to pay for their use of ikHerstel; operationalized as 1 self-reported item, scored using a 1-4 Likert scale
Habit[c]	The extent to which consumers tend to perform behaviors automatically because of learning	The degree to which patients feel their use of ikHerstel has become habitual; operationalized as 1 self-reported item, scored using a 1-4 Likert scale

[a]Based on the model by Steckler and Linnan [22].

[b]Not applicable.

[c]Based on the unified theory of acceptance and use of technology 2 by Venkatesh et al [28].

Data Analysis

Descriptive statistics were used to summarize the study's findings according to each process outcome as well as the UTAUT2 dimensions. Open-ended patient attitude items were assessed and categorized by the coordinating researcher, and the resultant categories were subsequently reviewed by another researcher from the research team.

Results

Reach

In the period between September 2022 and September 2023, 216 patients registered for the study. A schematic representation of the inclusion process is presented in Figure 2. Initial screening via telephone resulted in 148 exclusions. A major reason for exclusion was timing, as many patients only signed up for ikHerstel once their surgery had already taken place. The exclusion criteria were revised to account for this unexpected result, allowing patients to participate up to 14 days following their surgery. This nevertheless still led to 42 exclusions due to timing. A total of 68 patients were identified as eligible for participation and were subsequently sent informed consent forms. Among these 42 patients, 5 were excluded due to incompatible types of surgery that had not been identified as such prior to telephone screening. This resulted in a total of 63 included patients, which constitutes a reach of 81% (63 / (216 – (109 + 5 + 24)).

Baseline characteristics of these respondents are presented in Table 3. A majority of respondents were female, corresponding to one half of the included surgery types being gender specific for women. All the respondents had Dutch nationality and close to two-thirds (35/55) had a high level of education. All patients scored full points on the Quickscan test, and only one respondent gave a categorical self-description as being not very digitally skilled. The same held true for traditional literacy, with none of the respondents scoring in a range that would put them at risk of having low literacy skills [29].

Figure 2. Flow chart for inclusion in the study.

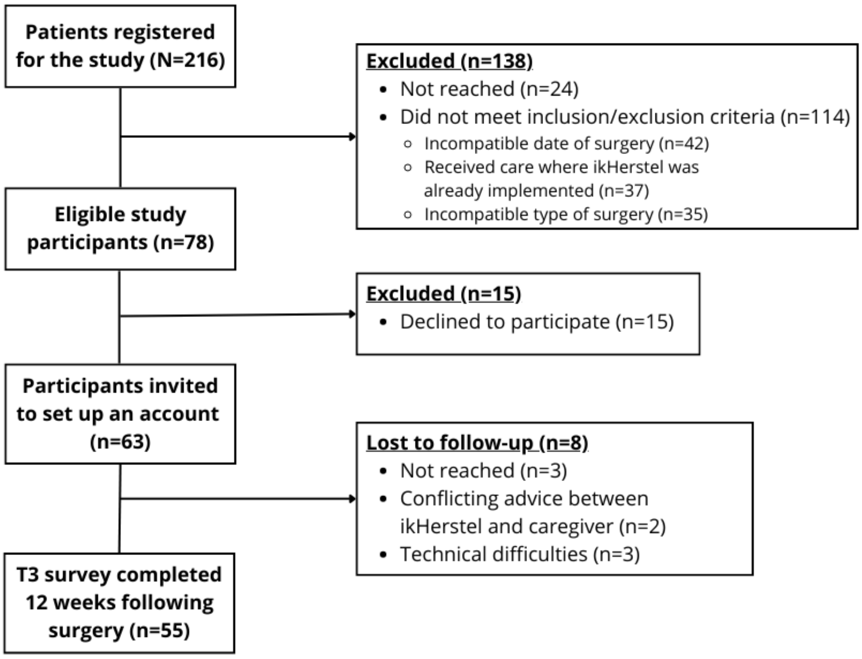


Table 3. Sample characteristics (n=55).

Variables	Values
Age (years), mean (SD)	48.6 (12.4)
Sex, n (%)	
Male	7 (13)
Female	48 (87)
Nationality, n (%)	
Dutch	55 (100)
Education, n (%)	
Low	7 (13)
Intermediate	13 (24)
High	35 (64)
Type of surgery, n (%)	
Laparoscopic uterus extirpation	21 (38)
Abdominal uterus extirpation	8 (15)
Vaginal uterus extirpation	6 (11)
Laparoscopic adnexal surgery	5 (9)
Laparoscopic cholecystectomy	10 (18)
Laparoscopic inguinal hernia surgery	4 (7)
Open inguinal hernia surgery	1 (2)
Digital skills—Quickscan, mean (SD)	6 (0)
Digital skills—self-scan (categorical), n (%)	
Very digitally skilled	29 (53)
Of average skill	25 (46)
Not or not very digitally skilled	1 (2)
Digital skills—self-scan (numeric), mean (SD)	7.9 (1.5)
Literacy score, mean (SD)	8.5 (2.6)

Dose Delivered

Of the 63 patients who signed the informed consent form and met the inclusion and exclusion criteria, 63 were provided with an account in the ikHerstel app. The dose-delivered fraction therefore computes to a percentage of 100%.

Dose Received

Of the 63 patients who were provided with an account, 55 activated their account and created a personalized recovery plan. Of these 55 patients, 34 reported using the app on a weekly or more frequent basis. The dose received fraction (34/63) therefore computes to a percentage of 54%.

Recruitment

An overview of the number of registrations and inclusions per recruitment tool is provided in Table 4. Most of the registrations (87/216, 40%) originated from tools that were dependent on hospitals, like posters, waiting room electronic displays, and hospital staff. Tools outside of the hospital yielded 36% (77/216) of registrations. However, they yielded more eligible patients (32 vs 31), as well as a higher proportion of eligible patients (32/77) compared to hospital-dependent tools (31/87).

Table 4. Overview of the number of registrations and eligible patients per recruitment tool.

Tools	Registrations, n (N=216)	Eligible patients, n (n=63)
Hospital-independent		
Forum advertisements	14	8
Webpage advertisements	1	0
Internet search	17	7
Social media	18	13
Magazine advertisements	21	2
Other[a]	6	2
Subtotal	77	32
Hospital-dependent		
Flyers	11	6
Posters	10	6
Business cards	2	2
Electronic displays	11	7
Hospital staff	24	8
Unspecified[b]	27	0
Other[c]	2	2
Subtotal	87	31
Unknown[d]	52	0

[a]This category included person-to-person contacts (n=5) and receiving an email of unknown origin (n=1).

[b]These respondents stated that the hospital was the source of their contact with ikHerstel.

[c]This category included patient-to-patient contacts in the convalescence room (n=1) and the webpage of the hospital (n=1).

[d]These respondents did not state how they came into contact with ikHerstel, mostly due to a lack of communication or stated interest on their part.

Patient Attitudes

Patients rated their overall satisfaction with ikHerstel an average 7.0 (SD 1.9) of 10. One patient did not answer the open-ended questions. A substantial proportion of patients (14/54) explicitly stated not having any dislikes about using ikHerstel, and an even greater proportion (29/54) explicitly reported no difficulties in using it. Most patients (49/54) reported positive experiences with ikHerstel. The most frequently stated (17/49) positive experience with ikHerstel related to its provision of perspective when it came to recovery. Patients furthermore found the app was clear in its presentation of information (10/49) and easy to use (8/46). Other stated likes related to the app's motivating power (6/49), its function as a source of information (3/49), its comforting effect (2/49), the patients' ability to benchmark their recovery (2/49), and a general statement of satisfaction (1/49). A majority of patients (50/54) reported on aspects that made using ikHerstel easy. The most frequently stated aspect was its clarity in presenting information (23/50). Patients also found it easy to navigate through the app (20/50) and praised its round-the-clock availability as a mobile phone app (6/50). One patient simply affirmed its ease of use, and others (4/50) found nothing about it easy. One patient stated, "Easy to use and provides motivation to start exercising and pick up activities again."

The most striking dislikes reported by patients were those concerning its recommendations. In some cases, what the app prescribed was misaligned with what patients felt they could handle. This mismatch ran both ways, as some patients felt the app was too ambitious, while others reported it was holding them back: "..that you [ikHerstel] go much faster than my recovery. That feels like failure because it repeatedly says you are behind on your recovery. It became more and more frustrating."

Another frequently stated mismatch was between ikHerstel and health care professionals. Of the 45 patients who reported receiving recovery recommendations from their health care professional, 33 stated that the recommendations provided by ikHerstel conflicted "sometimes" or more frequently. The majority of these (n=17) described the health care professional as conservative when it came to performing activities compared to the app. Others (n=8) reported that the app's recommendations were more elaborate and covered a wider slice of their daily life. Some patients (n=6) explicitly stated a dislike of the mismatch. In these cases as well, health care professionals' prescriptions were more conservative, and as a result, these patients reported feelings of frustration and insecurity: "[T]he recommendations from both the hospital and the GP [general practitioner]'s assistant were so much more conservative regarding when you should try and pick up activities that it made me feel insecure."

Other dislikes related to difficulties with inputting data (n=14), a lack of personalization (n=7), a lack of functionalities (n=5), the demotivating effect of the app (n=3), accessibility (n=1), technical failures (n=1), and miscellaneous difficulties (n=3); 14 patients found nothing to dislike. One patient stated, “After altering one of the activities, I had to redo all the input I had previously provided.”

UTAUT2 Dimensions

Among UTAUT2 survey dimensions, respondents rated their performance expectancy an average of 2.7 (SD 0.8) of 4 points. Effort expectancy was rated at 3.3 (SD 0.8) of 4 points and facilitating conditions at 3.4 (SD 0.7) of 4 points. The dimension of price value was scored an average 1.7 (SD 0.7) of 4 points, corresponding to 6 of 55 patients confirming that they would be agreeable to paying for the services provided by ikHerstel. A substantial proportion of patients (20/52) stated their use of ikHerstel had become habitual, resulting in an average score of 2.3 (SD 0.9) of 4 points for the dimension of habit.

Discussion

Principal Findings

In this feasibility study, we aimed to evaluate a patient-oriented implementation strategy for the mHealth intervention ikHerstel. We included 55 patients undergoing abdominal surgery among 216 registrations, and we investigated whether direct distribution of ikHerstel was a feasible addition to its implementation through hospitals. Hospital-dependent recruitment yielded slightly more registrations, while hospital-independent recruitment produced more eligible patients. The patient-oriented strategy constituted a reach of 81% (63/78), and 100% of reached patients were sent the intervention, after which 54% (34/63) engaged with it. Patients reported general satisfaction with ikHerstel, scoring it an average 7.0 (SD 1.9) of 10 points.

Other studies have examined user experiences with mHealth apps in the perioperative setting. To illustrate, a cross-sectional study on the Patient Journey app yielded higher levels of satisfaction compared to this study [15]. Patients were likewise positive about the app’s ease of use and its clear provision of useful information. A systematic review of patient experiences with mHealth confirms that this is a main benefit of these interventions [13]. The finding that patients regretted losing the possibility of communicating with their health care professional through the app was not replicated in our study. A previous process evaluation concerning a version of ikHerstel that did feature this function found that patients appreciated it, but that it should not replace a telephone appointment with their health care professional [24].

We hypothesized that the patient-oriented implementation strategy would increase ikHerstel’s reach. However, in terms of absolute scale, this expectation proved incorrect. Over the span of a year, only 216 registrations were generated, compared to the 1031 and 673 reported in previous studies, where hospitals played a central role in recruitment through their waiting lists [24,27]. Despite lower registration numbers, the reach of the patient-oriented implementation strategy was better, or at least comparable to, previous studies, at 81%, compared to 40% and 60%, respectively [24,27]. In addition to scale, an advantage of recruitment through hospitals was apparent when comparing the rate of and reasons for exclusion. Only 5% of patients were excluded due to ineligibility in the study by van der Meij et al [24], compared to our study’s exclusion rate of 53%. Poor timing (n=42, 37%), double registration (n=37, 32%), and ineligible types of surgery (n=35, 31%) make up the reasons for exclusion. In fact, poor timing proved such a barrier to participation that we were forced to revise our exclusion criteria halfway through the study to include patients up to 14 days after their surgery. Our assumption that patients would start looking for tools to support them through their perioperative journey prior to surgery proved false. In practice, this means that a substantial proportion of patients missed out on ikHerstel’s preoperative functions designed to enhance preparation and manage expectations.

The mismatch between ikHerstel’s recommendations and those of health care professionals also points to the strategic position of these professionals in perioperative care. Patients listed this mismatch not only as a source of dislike but also as one of feelings of insecurity. Other studies have reported similar findings [13,15]. The conflict itself may arise due to the conservative character of many health care professionals, as some studies indicate [30,31]. Complications that arose may likewise have caused mismatches by altering patients’ needs and invalidating the care provision of ikHerstel. Both cases advocate for the integral role of health care professionals in mHealth implementation strategies, as they are ideally situated to select patients and to adjust care provision when complications arise. By replacing these agents with a researcher, we effectively placed a part of our intervention outside of the broader system of care. Despite this, most patients had no trouble using ikHerstel independently. More than half of patients reported no difficulties and a quarter of patients explicitly found nothing to dislike.

Patients find value in mHealth apps in their provision of information that would otherwise not be readily available, and find even more value if that information is tailored to the patients’ individual situation [32]. In light of our own findings, it seems vital that health care professionals are involved in how mHealth is implemented to provide this function: as gatekeepers, selecting the right patients; as anchors, integrating an intervention into the broader system of care; but not as tech support, as patients seem able to navigate mHealth independently. Health care professionals could be involved through professional training, introducing them to the mHealth evidence base, or it may take the form of colleagues operating as implementation champions [33].

Limitations

A number of limitations need to be addressed, the first being the absence of health care professionals’ perspectives in our evaluation of the implementation strategy’s feasibility. The patient-oriented character of the study was chosen in dialogue with patient interest groups and health insurers, and aligns with the study’s aim of empowering patients to access ikHerstel even if their hospital has not implemented it. Health care professionals’ assessments of our strategy may nevertheless have yielded important insights, as they may have shed light on conflicting recovery recommendations that were received by the participants.

Another limitation is the study's lack of a diverse sample of patients. We disproportionately included highly educated women of Dutch nationality. While an overrepresentation of women was expected due to the overrepresentation of gynecological types of surgery in our study, this does not explain the sample's high level of education or the lack of international patients. In the case of the latter, the use of the Dutch language in our recruitment material may well have discouraged any international patients from engaging with the study. For the former, the multimedia recruitment strategy we used, emphasizing access to a medical innovation, may have selected for highly educated patients, as some studies have reported on the association between educational level and the use of health services [34-37]. Here too, we may see a reflection of the absence of a health care professional, whose prompting influence might have worked to transcend such barriers. A study on sex differences regarding intention to use mHealth apps in the Netherlands found that women had a more negative attitude of mHealth, perceiving it as being less useful than did men [38]. This may have driven the difference in overall satisfaction scores between this study and the previous study by van der Meij et al [24], who included a more equal distribution of male versus female patients. Stratification by sex provides some weight to this argument, producing an average satisfaction score of 8.3 for men versus 6.8 for women, although these figures lack reliability precisely due to our sample's low representation of men.

Conclusions

The patient-oriented implementation strategy evaluated in this study was an equivocal success. One of its main hypothesized advantages of more easily reaching a wide audience of patients was not demonstrated. However, its method of recruitment has low costs, and most patients were satisfied and engaged with the mHealth app. Lack of involvement of health care professionals, rather than usability issues on the patients' side, contributed to patients' concerns regarding conflicting recommendations between ikHerstel and health care professionals. Given patient engagement, satisfaction, and improvement in outcomes [12,16-18] with use of such apps, hospitals should consider strategies where health care professionals are involved in selecting patients that may benefit from mHealth apps for postoperative recovery after day surgery and guiding patients' care.

Acknowledgments

We would like to thank Jeroen de Wilde, Tim Thurlings, Patiëntenfederatie Nederland, and Coöperatie VGZ for helping collect patient data.

Conflicts of Interest

EvdM, JAFH, and JRA are the developers of the mHealth care program under study. JAFH and JRA are consultants and certificate holders of a spinoff company for implementation of the mobile app component of the IkHerstel intervention in the Netherlands (ie, the intervention under study). This spinoff company had no impact on the submitted work. JAFH received grants from Nederlandse Organisatie voor Wetenschappelijk Onderzoek, ZonMw, and Samsung during the conduct of the study and received a fee from Olympus outside the submitted work. JRA holds a chair in insurance medicine paid by the Dutch Social Security Agency and has received grants from ZonMw, Nederlandse Organisatie voor Wetenschappelijk Onderzoek, Instituut Gak, Uitvoeringsinstituut Werknemersverzekeringen, Sociale Zaken en Werkgelegenheid, VWS (Volksgezondheid, Welzijn en Sport), Pfizer, Achmea, CVZ (College Voor Zorgverzekeringen), and Zorginstituut; all outside the submitted work. EvdM declares no competing interests.

Multimedia Appendix 1
Screening structure and content of ikHerstel.
[DOCX File , 316 KB - periop_v8i1e58878_app1.docx]

Multimedia Appendix 2
Advertisement design.
[DOCX File , 796 KB - periop_v8i1e58878_app2.docx]

Multimedia Appendix 3
Unified theory of acceptance and use of technology survey items.
[DOCX File , 22 KB - periop_v8i1e58878_app3.docx]

References

1. Health at a glance: Europe. Organisation for Economic Co-operation and Development/European Union. 2018. URL: https://www.oecd.org/en/publications/health-at-a-glance-europe-2018_health_glance_eur-2018-en/full-report.html [accessed 2025-01-07]

2. Madsen H, Henderson W, Dyas A, Bronsert M, Colborn K, Lambert-Kerzner A, et al. Inpatient versus outpatient surgery: a comparison of postoperative mortality and morbidity in elective operations. World J Surg 2023 Mar;47(3):627-639. [doi: 10.1007/s00268-022-06819-z] [Medline: 36380104]
3. Friedlander DF, Krimphove MJ, Cole AP, Marchese M, Lipsitz SR, Weissman JS, et al. Where is the value in ambulatory versus inpatient surgery? Ann Surg 2021 May 01;273(5):909-916. [doi: 10.1097/SLA.0000000000003578] [Medline: 31460878]
4. Lemos P, Pinto A, Morais G, Pereira J, Loureiro R, Teixeira S, et al. Patient satisfaction following day surgery. J Clin Anesth 2009 May;21(3):200-205. [doi: 10.1016/j.jclinane.2008.08.016] [Medline: 19464614]
5. Martin AD, Nunez RN, Andrews JR, Martin GL, Andrews PE, Castle EP. Outpatient prostatectomy: too much too soon or just what the patient ordered. Urology 2010 Feb;75(2):421-424. [doi: 10.1016/j.urology.2009.08.085] [Medline: 19969327]
6. Mathis MR, Naughton NN, Shanks AM, Freundlich RE, Pannucci CJ, Chu Y, et al. Patient selection for day case-eligible surgery: identifying those at high risk for major complications. Anesthesiology 2013 Dec;119(6):1310-1321 [FREE Full text] [doi: 10.1097/ALN.0000000000000005] [Medline: 24108100]
7. Tran TT, Kaneva P, Mayo NE, Fried GM, Feldman LS. Short-stay surgery: what really happens after discharge? Surgery 2014 Jul;156(1):20-27. [doi: 10.1016/j.surg.2014.03.024] [Medline: 24856316]
8. Berg K, Arestedt K, Kjellgren K. Postoperative recovery from the perspective of day surgery patients: a phenomenographic study. Int J Nurs Stud 2013 Dec;50(12):1630-1638. [doi: 10.1016/j.ijnurstu.2013.05.002] [Medline: 23726224]
9. Gilmartin J. Contemporary day surgery: patients' experience of discharge and recovery. J Clin Nurs 2007 Jun;16(6):1109-1117. [doi: 10.1111/j.1365-2702.2007.01548.x] [Medline: 17518885]
10. Larsson F, Strömbäck U, Rysst Gustafsson S, Engström Å. Postoperative recovery: experiences of patients who have undergone orthopedic day surgery. J Perianesth Nurs 2022 Aug;37(4):515-520 [FREE Full text] [doi: 10.1016/j.jopan.2021.10.012] [Medline: 35279387]
11. Mottram A. 'They are marvellous with you whilst you are in but the aftercare is rubbish': a grounded theory study of patients' and their carers' experiences after discharge following day surgery. J Clin Nurs 2011 Nov;20(21-22):3143-3151. [doi: 10.1111/j.1365-2702.2011.03763.x] [Medline: 21762418]
12. Vonk Noordegraaf A, Anema JR, van Mechelen W, Knol DL, van Baal WM, van Kesteren PJM, et al. A personalised eHealth programme reduces the duration until return to work after gynaecological surgery: results of a multicentre randomised trial. BJOG 2014 Aug;121(9):1127-35; discussion 1136. [doi: 10.1111/1471-0528.12661] [Medline: 24511914]
13. De La Cruz Monroy MFI, Mosahebi A. The use of smartphone applications (apps) for enhancing communication with surgical patients: a systematic review of the literature. Surg Innov 2019 Apr;26(2):244-259. [doi: 10.1177/1553350618819517] [Medline: 30602332]
14. Timmers T, Janssen L, Kool RB, Kremer JA. Educating patients by providing timely information using smartphone and tablet apps: systematic review. J Med Internet Res 2020 Apr 13;22(4):e17342 [FREE Full text] [doi: 10.2196/17342] [Medline: 32281936]
15. Willems SJ, Coppieters MW, Pronk Y, Diks MJF, van der Heijden KWAP, Rooker S, et al. A clinical journey mobile health app for perioperative patients: cross-sectional study. JMIR Hum Factors 2021 Feb 08;8(1):e20694 [FREE Full text] [doi: 10.2196/20694] [Medline: 33555262]
16. den Bakker CM, Schaafsma FG, Consten ECJ, Schraffordt Koops SE, van der Meij E, van de Ven PM, et al. Personalised electronic health programme for recovery after major abdominal surgery: a multicentre, single-blind, randomised, placebo-controlled trial. Lancet Digit Health 2023 Aug;5(8):e485-e494 [FREE Full text] [doi: 10.1016/S2589-7500(23)00084-5] [Medline: 37419843]
17. Bouwsma EVA, Bosmans JE, van Dongen JM, Brölmann HAM, Anema JR, Huirne JAF. Cost-effectiveness of an internet-based perioperative care programme to enhance postoperative recovery in gynaecological patients: economic evaluation alongside a stepped-wedge cluster-randomised trial. BMJ Open 2018 Jan 21;8(1):e017782 [FREE Full text] [doi: 10.1136/bmjopen-2017-017782] [Medline: 29358423]
18. van der Meij E, Anema J, Leclercq W, Bongers M, Consten E, Schraffordt Koops S, et al. Personalised perioperative care by e-health after intermediate-grade abdominal surgery: a multicentre, single-blind, randomised, placebo-controlled trial. Lancet 2018 Jul 07;392(10141):51-59. [doi: 10.1016/S0140-6736(18)31113-9] [Medline: 29937195]
19. Ha JF, Longnecker N. Doctor-patient communication: a review. Ochsner J 2010;10(1):38-43 [FREE Full text] [Medline: 21603354]
20. Kvarnström K, Westerholm A, Airaksinen M, Liira H. Factors contributing to medication adherence in patients with a chronic condition: a scoping review of qualitative research. Pharmaceutics 2021 Jul 20;13(7):1100 [FREE Full text] [doi: 10.3390/pharmaceutics13071100] [Medline: 34371791]
21. Praatplaat belemmeringen opschaling telemonitoring. Nederlandse Vereniging van Ziekenhuizen. URL: https://nvz-ziekenhuizen.nl/sites/default/files/2022-12/Belemmeringen%20opschaling%20telemonitoring.pdf [accessed 2024-02-19]
22. Steckler A, Linnan L. Process Evaluation for Public Health Interventions and Research. Hoboken, NJ: Jossey-Bass; 2002.
23. den Bakker CM, Schaafsma FG, van der Meij E, Meijerink WJ, van den Heuvel B, Baan AH, et al. Electronic health program to empower patients in returning to normal activities after general surgical and gynecological procedures: intervention

mapping as a useful method for further development. J Med Internet Res 2019 Feb 06;21(2):e9938 [FREE Full text] [doi: 10.2196/jmir.9938] [Medline: 30724740]

24. van der Meij E, Huirne JA, Ten Cate AD, Stockmann HB, Scholten PC, Davids PH, et al. A perioperative eHealth program to enhance postoperative recovery after abdominal surgery: process evaluation of a randomized controlled trial. J Med Internet Res 2018 Jan 02;20(1):e1 [FREE Full text] [doi: 10.2196/jmir.8338] [Medline: 29295808]
25. Greef M, Deursen A, Tubbing M, Bohnenn E. Development of the Diagnostic Illiteracy Scale in order to reveal illiteracy among adults. Andragogic Stud J Stu Adult Edu Learn 2013;1(1):1-48 [FREE Full text]
26. Quickscan digitale vaardigheden. Pharos. URL: https://www.pharos.nl/wp-content/uploads/2024/08/Quickscan_digitale-vaardigheden_082024.pdf [accessed 2022-05-02]
27. Bouwsma EVA, Vonk Noordegraaf A, Szlávik Z, Brölmann HAM, Emanuel MH, Lips JP, et al. Process evaluation of a multidisciplinary care program for patients undergoing gynaecological surgery. J Occup Rehabil 2014 Sep;24(3):425-438 [FREE Full text] [doi: 10.1007/s10926-013-9475-4] [Medline: 24057871]
28. Venkatesh V, Thong JYL, Xu X. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Q 2012;36(1):157-178. [doi: 10.2307/41410412]
29. van Deursen A, van Dijk J. Modeling traditional literacy, internet skills and internet usage: an empirical study. Interact Comput 2014 Jul 16;28(1):13-26. [doi: 10.1093/iwc/iwu027]
30. Clayton M, Verow P. Advice given to patients about return to work and driving following surgery. Occup Med (Lond) 2007 Oct;57(7):488-491. [doi: 10.1093/occmed/kqm063] [Medline: 17906266]
31. Nygaard IE, Hamad NM, Shaw JM. Activity restrictions after gynecologic surgery: is there evidence? Int Urogynecol J 2013 May;24(5):719-724 [FREE Full text] [doi: 10.1007/s00192-012-2026-2] [Medline: 23340879]
32. Peng W, Kanthawala S, Yuan S, Hussain SA. A qualitative study of user perceptions of mobile health apps. BMC Public Health 2016 Nov 14;16(1):1158 [FREE Full text] [doi: 10.1186/s12889-016-3808-0] [Medline: 27842533]
33. Santos WJ, Graham ID, Lalonde M, Demery Varin M, Squires JE. The effectiveness of champions in implementing innovations in health care: a systematic review. Implement Sci Commun 2022 Jul 22;3(1):80 [FREE Full text] [doi: 10.1186/s43058-022-00315-0] [Medline: 35869516]
34. Groeneveld PW, Sonnad SS, Lee AK, Asch DA, Shea JE. Racial differences in attitudes toward innovative medical technology. J Gen Intern Med 2006 Jun;21(6):559-563 [FREE Full text] [doi: 10.1111/j.1525-1497.2006.00453.x] [Medline: 16808736]
35. National Academies of Sciences, Engineering, and Medicine. Health-Care Utilization as a Proxy in Disability Determination. Washington, DC: The National Academies Press; 2018.
36. Salganicoff A, Ranji U, Beamesderfer A, Kurani N. Women and health care in the early years of the ACA: key findings from the 2013 Kaiser Women's Health Survey. KFF. 2014 May 15. URL: https://www.kff.org/womens-health-policy/report/women-and-health-care-in-the-early-years-of-the-aca-key-findings-from-the-2013-kaiser-womens-health-survey/ [accessed 2025-07-01]
37. Hopman P, Heins M, Rijken M, Schellevis F. Health care utilization of patients with multiple chronic diseases in the Netherlands: differences and underlying factors. Eur J Intern Med 2015 Apr;26(3):190-196 [FREE Full text] [doi: 10.1016/j.ejim.2015.02.006] [Medline: 25704328]
38. van Elburg FRT, Klaver NS, Nieboer AP, Askari M. Gender differences regarding intention to use mHealth applications in the Dutch elderly population: a cross-sectional study. BMC Geriatr 2022 May 24;22(1):449 [FREE Full text] [doi: 10.1186/s12877-022-03130-3] [Medline: 35610577]

Abbreviations

UTAUT2: unified theory of acceptance and use of technology 2

Edited by N Rohatgi; submitted 27.03.24; peer-reviewed by H Karim, A Santiago, D Poenaru, M van der Velde, A Lesher; comments to author 11.09.24; revised version received 06.12.24; accepted 16.12.24; published 14.01.25.

Please cite as:
Toben D, de Wind A, van der Meij E, Huirne JAF, Anema JR
A Patient-Oriented Implementation Strategy for a Perioperative mHealth Intervention: Feasibility Cohort Study
JMIR Perioper Med 2025;8:e58878
URL: https://periop.jmir.org/2025/1/e58878
doi:10.2196/58878
PMID:

Original Paper

Development and Validation of a Routine Electronic Health Record-Based Delirium Prediction Model for Surgical Patients Without Dementia: Retrospective Case-Control Study

Emma Holler[1], PhD; Christina Ludema[2], PhD; Zina Ben Miled[3], PhD; Molly Rosenberg[2], PhD; Corey Kalbaugh[2], PhD; Malaz Boustani[4], MD; Sanjay Mohanty[1], MD

[1]Department of Surgery, Indiana University School of Medicine, Indianapolis, IN, United States

[2]Department of Epidemiology & Biostatistics, Indiana University Bloomington, Bloomington, United States

[3]Department of Electrical & Computer Engineering, Lamar University, Beaumont, TX, United States

[4]Department of Medicine, Indiana University School of Medicine, Indianapolis, IN, United States

Corresponding Author:
Emma Holler, PhD
Department of Surgery
Indiana University School of Medicine
545 Barnhill Drive
Indianapolis, IN, 46202
United States
Phone: 1 317 944 5376
Email: emorone@iu.edu

Abstract

Background: Postoperative delirium (POD) is a common complication after major surgery and is associated with poor outcomes in older adults. Early identification of patients at high risk of POD can enable targeted prevention efforts. However, existing POD prediction models require inpatient data collected during the hospital stay, which delays predictions and limits scalability.

Objective: This study aimed to develop and externally validate a machine learning-based prediction model for POD using routine electronic health record (EHR) data.

Methods: We identified all surgical encounters from 2014 to 2021 for patients aged 50 years and older who underwent an operation requiring general anesthesia, with a length of stay of at least 1 day at 3 Indiana hospitals. Patients with preexisting dementia or mild cognitive impairment were excluded. POD was identified using Confusion Assessment Method records and delirium International Classification of Diseases (ICD) codes. Controls without delirium or nurse-documented confusion were matched to cases by age, sex, race, and year of admission. We trained logistic regression, random forest, extreme gradient boosting (XGB), and neural network models to predict POD using 143 features derived from routine EHR data available at the time of hospital admission. Separate models were developed for each hospital using surveillance periods of 3 months, 6 months, and 1 year before admission. Model performance was evaluated using the area under the receiver operating characteristic curve (AUROC). Each model was internally validated using holdout data and externally validated using data from the other 2 hospitals. Calibration was assessed using calibration curves.

Results: The study cohort included 7167 delirium cases and 7167 matched controls. XGB outperformed all other classifiers. AUROCs were highest for XGB models trained on 12 months of preadmission data. The best-performing XGB model achieved a mean AUROC of 0.79 (SD 0.01) on the holdout set, which decreased to 0.69-0.74 (SD 0.02) when externally validated on data from other hospitals.

Conclusions: Our routine EHR-based POD prediction models demonstrated good predictive ability using a limited set of preadmission and surgical variables, though their generalizability was limited. The proposed models could be used as a scalable, automated screening tool to identify patients at high risk of POD at the time of hospital admission.

(JMIR Perioper Med 2025;8:e59422) doi:10.2196/59422

KEYWORDS

delirium; machine learning; prediction; postoperative; algorithm; electronic health records; surgery; risk prediction

Introduction

Postoperative delirium (POD) is a common and serious surgical complication that affects 15%-50% of older surgical patients [1-3]. POD is characterized by acute fluctuations in consciousness and has a complex etiology thought to be caused by interactions between predisposing (eg, individual vulnerability) and precipitating (eg, acute illness or surgery) factors [4]. Common predisposing factors include older age, preexisting cognitive impairment, poor physical functioning, alcohol abuse, smoking, and depression [5-8]. Risk factors unique to surgical settings include the type of surgery (eg, major vascular procedures), emergent status, case complexity, and perioperative medications [6,7,9,10]. Despite being an acute condition, delirium is associated with long-term cognitive and physical impairment, institutionalization, and death [4,11]. However, up to 40% of cases may be preventable, and multicomponent, nonpharmacologic interventions may be effective in reducing incidence and health care costs [12,13].

Early and accurate POD risk prediction can inform prevention and enable targeted intervention and resource planning efforts. Fortunately, the widespread availability of electronic health record (EHR) data and advancements in machine learning offer an opportunity to develop accurate, low-cost, and scalable screening tools for POD risk. Several machine learning-based POD prediction models have been developed, reporting areas under the curve (AUROCs) ranging from 0.71 to 0.86 [14-26]. However, the models with the highest AUROCs have important limitations that hinder their practical application. First, they focus on specific patient subsets (ie, intensive care unit (ICU) patients, cardiac surgery), which restricts their generalizability to general surgical populations. Second, population-specific models necessitate separate models for each subpopulation, making implementation cumbersome and resource intensive. Finally, many of these models require inpatient data that take hours or days to accumulate, delaying risk assessment and potential interventions. A small number of studies have developed POD prediction models for general surgical populations; however, these models still incorporate nonroutine clinical data (eg, inpatient nursing assessments) that require time to collect and may not be universally available [14-18,27].

These limitations highlight the need for a model that can predict POD in a diverse surgical population using readily available preoperative data, as it could provide an early, inexpensive, and scalable prescreening tool to identify patients who may benefit from additional monitoring or preventative measures. In this study, we developed and externally validated a machine learning model that can accurately predict POD in surgical patients at the time of hospital admission using only routine EHR data. We also identified preoperative EHR-based predictors of POD and determined how preoperative surveillance length affected model performance.

Methods

Ethical Considerations

This study was approved by the Indiana University (IU) Institutional Review Board (#15767) and adhered to the reporting standards described in the Transparent Reporting of Individual Prognosis or Diagnosis (TRIPOD) guidelines [27].

Study Data and Cohort Selection

Diagnoses, medication orders, surgery, and other inpatient clinical records (eg, nursing assessments) were extracted from the IU Health electronic data warehouse. IU Health, a nonprofit health system with the largest physician network in the state of Indiana, includes 17 hospitals and dozens of outpatient facilities and performs approximately 115,000 surgeries per year [28]. We identified all surgical hospitalizations for patients aged 50 years and older who underwent surgery requiring general anesthesia at an IU Health facility between January 1, 2014, and December 31, 2021; had a length of stay of at least 1 day; and did not have preexisting dementia. Hospitalizations of patients with preexisting dementia (defined as having a dementia diagnosis code or an order for an antidementia medication before admission; see Table S1 in Multimedia Appendix 1) were excluded because dementia is known to be the single-strongest predictor of delirium [6]; models are not needed to forecast risk. For a hospitalization to be eligible, the patient had to have at least 1 IU Health encounter (defined as any interaction with an IU Health facility, eg, outpatient, inpatient, or emergency department visits) in the year before admission and have at least 1 diagnosis or medication record during that period. If no sex, race, or age data were available across all of a given patient's hospitalizations, that patient was excluded.

This study followed a retrospective case-control design where nondelirium (ie, control) hospitalizations were matched to delirium (ie, case) hospitalizations by sex, race, age within 3 years, and admission year within 3 years. We matched on these variables to ensure the age distribution for cases and controls was equalized across race and sex groups. As a result, age was less important to the model, and biases within strata of race and sex were minimized. Because matching was done at the hospitalization level rather than the patient-level, it was possible for case and control hospitalizations belonging to the same patient to be matched.

Hospitalizations where the patient developed POD were designated as cases. POD was defined as at least 1 positive Confusion Assessment Method (CAM) [29] nursing assessment or a delirium *International Classification of Diseases, Ninth Revision* (ICD-9)/*International Classification of Diseases, Tenth Revision, Clinical Modification* (ICD-10-CM) code (see Table S2 in Multimedia Appendix 1) recorded during the hospital stay. The CAM is a validated diagnostic algorithm with an overall sensitivity of 94% and a specificity of 89% [30]. Hospitalizations where delirium was present at the time of admission were excluded because the model is intended to predict POD. Hospitalizations without delirium or any nurse-documented confusion (ie, cognitive assessments reporting that the patient was disoriented, confused, or did not follow commands) were eligible to be selected as controls. Visits that did not have documented delirium (ie, delirium ICD code or positive CAM) but did have nurse-documented confusion were excluded from the control pool to ensure controls were not actually misclassified cases; confusion (without delirium) could possibly represent subsyndromal delirium. If a case had more than 1 potential control, a control was randomly selected. For each eligible visit, the index date was defined as the date of hospital admission. We used the following set of sociodemographic, surgery, diagnosis, and medication variables to build our predictive models.

Variables

Sociodemographic variables included age, patient-reported sex, and patient-reported race (categorized as Black, White, Asian, other, or unknown for analytic purposes), and insurance type. The insurance type was ascertained during each index visit and categorized as commercial, government (Medicare or Medicaid), self-pay, or other/unknown. Smoking status at the time of surgery was extracted from the EHR and categorized as "current," "former," or "never smoker." The BMI was obtained from the visit nearest to the index. The initial American Society of Anesthesiologists (ASA) class and emergency surgery status (defined as operations with an ASA class of 5 or E) were also included. Surgical specialty was assigned based on National Surgical Quality Improvement Program inclusion and exclusion criteria [31]**.** If a patient underwent 4 or more procedures falling under 2 or more distinct specialties, the visit was categorized as "multispecialty."

Diagnosis variables were generated using ICD-9/ICD-10-CM codes. Binary variables were created for each of the 31 Elixhauser disease groups using Quan et al [32] coding scheme and Elixhauser mortality scores were calculated for each patient using van Walravan weights [32-34]. We also created binary variables for other diagnoses potentially associated with increased risk of delirium, including previous delirium, cerebrovascular disease (CVD), previous traumatic brain injury (TBI), and sensory impairment (Table S3 in Multimedia Appendix 1). We derived a composite variable representing the total comorbidity burden by calculating the sum of the number of unique ICD codes (at the 3-digit level) a patient had prior to each index date. Variables for the number of ICD codes belonging to the ICD-10 group Z00-Z99 (factors influencing health status and contact with health services) and their ICD-9 equivalents were also included based on prior literature [14], grouped as follows: Z00-Z13, Z16, Z17, Z18, Z20-29, Z30-39, Z40-53, Z55-65, Z69-76, and Z77-99.

Medication variables were generated using medication order data. Anticholinergic (ACh) medications were identified using the Anticholinergic Cognitive Burden (ACB) scale, a well-established tool that categorizes medications based on the strength of their ACh activity [35]. Three ACh medication variables were developed representing the total number of orders for drugs with an ACB score of 1, 2, and 3, respectively. We also included other non-ACh medication variables as predictors. Since medication orders were retrieved from multiple health care institutions, a unified mapping of medication names to a drug taxonomy was not available. Instead, we mapped each medication in the medication orders to the Anatomical Therapeutic Chemical (ATC) classification codes [36]. The ATC drug classification system is hierarchical with multiple sublevels and maintained by the World Health Organization. For this study, all 14 main groups (eg, A: alimentary tract and metabolism; B: blood and blood-forming organs; C: cardiovascular system) and the first-level subgroup were included (eg, A01: stomatological preparations; A02: drugs for acid-related disorders). For each patient, the count of medication orders (excluding AChs, which were derived separately, as described before) associated with a given ATC subgroup was calculated over the preindex assessment period. We also created a variable summing the total number of medication orders before each admission to capture polypharmacy.

Model Development and Evaluation

Three IU Health institutions were selected for this study. Institutions A, B, and C had the first-, second-, and third-greatest number of delirium cases, respectively. Institution-specific models were developed using data derived from the following preindex surveillance periods: 3 months before admission, 6 months before admission, and 1 year before admission. The purpose of training these separate models was to provide an understanding of how the training data and surveillance period impact the models' ability to predict POD and generalizability. Prior to training, each model's data were split into training (80%) and holdout (20%) sets, while maintaining a 1:1 ratio of cases and controls to avoid class imbalance. Imbalanced data are problematic in classification tasks because the model will focus on learning the characteristics of the majority class. As a result, the model may achieve high accuracy but fail to accurately identify the minority class.

In this study, 6 demographic variables, 4 surgical variables, 49 diagnosis variables, and 84 medication variables were included for a total of 143 features. Categorical variables were one-hot encoded (ie, converted into dummy variables), and continuous variables were standardized such that they each had a mean of 0 and an SD of 1. We initially explored several different machine learning models to predict whether patients would develop POD after surgery. In addition to traditional logistic regression, a parametric model, we also tried random forest, extreme gradient boosting (XGB), and a multilayer neural network because they can learn complex nonlinear relationships between variables. Optimal hyperparameters for each model were selected using a grid search with 5-fold cross-validation. Each candidate model was evaluated by calculating the area under the receiver operating characteristic curve (AUROC) on its holdout set using data from 1 year before hospital admission, and the model with the highest AUROC was selected as the final model. XGB outperformed the other candidate classifiers in all cases.

After model selection, XGB models trained on data from institution A (referred to as XGB_A) were internally validated on holdout data from institution A and externally validated using holdout data from institutions B and C. Similarly, models trained on data from institutions B and C (referred to as XGB_B and XGB_C, respectively) were internally validated on holdout data from institutions B and C and externally validated using data from institutions A and C and A and B, respectively. The predictive performance of each model was evaluated on the holdout and external validation data by creating 1000 bootstrapped samples without replacement, calculating the AUROC, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) in each sample and then averaging them across all samples. We also generated predictions for nondelirium visits with nurse-documented confusion (which were excluded from training) to examine how the models handle patients with possible subsyndromal delirium. The default threshold of 0.50 was used for predictions. Shapley Additive Explanation (SHAP) [37] was used to determine the most important features, and model calibration was assessed using calibration curves. All analyses were completed using R version 4.3.2 (R Foundation for Statistical Computing).

Results

Study Cohort

Figure 1 depicts the workflow used for model development, internal validation, and external validation for the model trained on data from institution A. Between 2014 and 2022, at the 3 institutions of interest, there were 39,968 surgical visits for 30,131 unique patients aged 50 years and older. Of the identified visits, 431 (1.4%) were excluded for not having any previous diagnosis or medication order data, and 120 (0.4%) were excluded for missing sex, race, or the ASA class. The 6250 (20.7%) visits with nurse-documented confusion (but no delirium) were excluded from the training and holdout sets but reserved for later analyses. After matching, the final analytic sample included 7167 (23.8%) delirium cases and 7167 (23.8%) matched controls (Figure 2).

Figure 1. Workflow for the development and validation of the model using data from institution A. XGB: extreme gradient boosting.

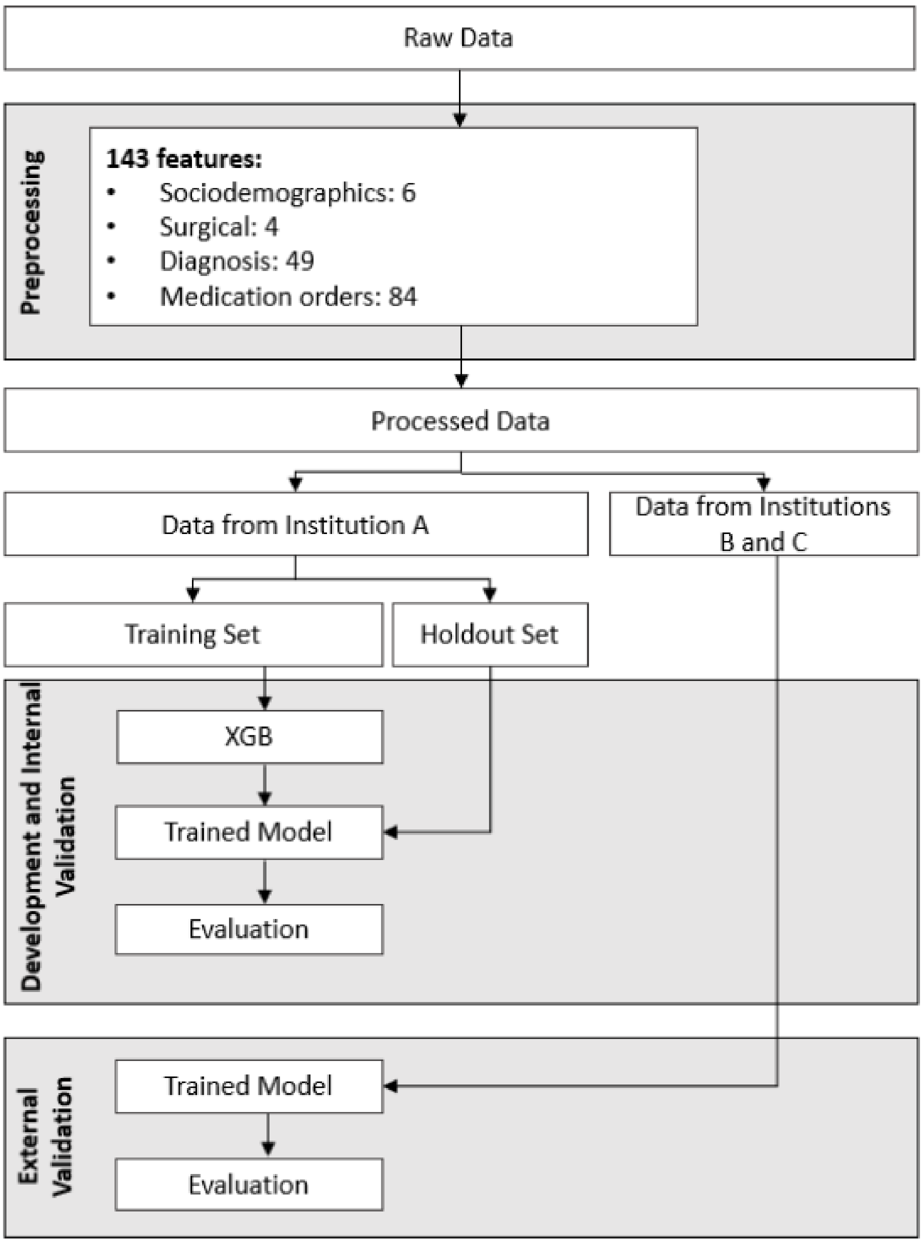
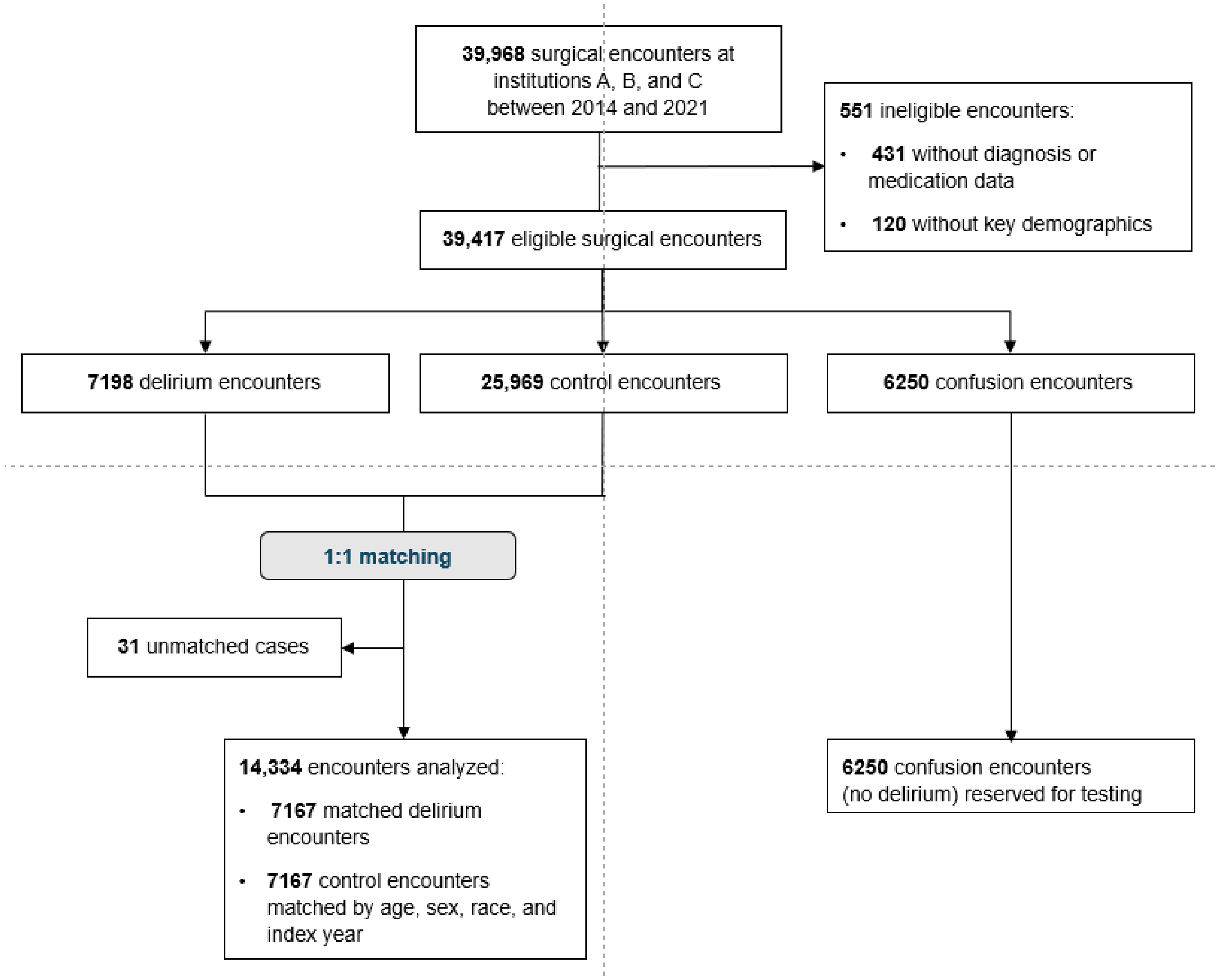


Figure 2. Patient inclusion flow diagram.



Pooling across institutions, the median age was 68 (IQR 61-76) years, and most patients were male (n=7412, 51.7%), White (n=12,276, 85.6%), and had public insurance (n=11,523, 80.4%). The most common surgical specialty was general surgery (n=3600, 25.1%), and 11.5% (n=1644) of operations were classified as emergencies (Table 1 and Table S4 in Multimedia Appendix 1).

As shown in Table 2, the 3 most common comorbidities in the general cohort were hypertension (n=9998, 69.8%), diabetes (n=5189, 36.2%), and nonmetastatic cancer (n=5222, 29.6%). Delirium cases differed from controls in several respects. Delirium cases had a greater comorbidity burden than controls and were more likely to have previous delirium (Table 2 and Table S5 in Multimedia Appendix 1).

Of the 6250 (20.7%) visits with nurse-documented confusion but without delirium, 3185 (51%) belonged to institution A, 1328 (21.2%) to institution B, and 1737 (27.8%) to institution C. Patients with confusion were more likely to have had delirium in the past year than controls but less likely than cases. Their comorbidity burden also fell in between that of cases and controls (Tables S6 and S7 in Multimedia Appendix 1).

Table 1. Characteristics of delirium cases and controls by institution.

Variables[a]	Institution A		Institution B		Institution C	
	Controls (n=3739)	Cases (n=3739)	Controls (n=1928)	Cases (n=1928)	Controls (n=1500)	Cases (n=1500)
Age (years), median (IQR)	68 (61-76)	68 (61-76)	66 (59-73)	66 (59-73)	72 (63-80)	72 (63-80)
Sex: female, n (%)	1840 (49.2)	1840 (49.2)	861 (44.7)	861 (44.7)	760 (50.7)	760 (50.7)
Race, n (%)						
Asian	12 (0.3)	12 (0.3)	13 (0.7)	13 (0.7)	1 (0.1)	1 (0.1)
Black	758 (20.3)	758 (20.3)	162 (8.4)	162 (8.4)	59 (3.9)	59 (3.9)
Other	4 (0.1)	4 (0.1)	3 (0.2)	3 (0.2)	4 (0.3)	4 (0.3)
White	2959 (79.1)	2959 (79.1)	1747 (90.6)	1747 (90.6)	1432 (95.5)	1432 (95.5)
Unknown	6 (0.2)	6 (0.2)	3 (0.2)	3 (0.2)	4 (0.3)	4 (0.3)
Insurance, n (%)						
Private	857 (22.9)	572 (15.3)	547 (28.4)	391 (20.3)	239 (15.9)	124 (8.3)
Public	2861 (76.5)	3137 (83.9)	1376 (71.4)	1530 (79.4)	1253 (83.5)	1366 (91.1)
Uninsured	21 (0.6)	30 (0.8)	5 (0.3)	7 (0.4)	8 (0.5)	10 (0.7)
BMI, median (IQR)	28.5 (24.3-33.7)	27.5 (23.1-32.7)	27.2 (23.2-32.0)	27.0 (22.7-32.0)	28.0 (23.9-33.6)	27.2 (22.9-33.2)
Smoking status, n (%)						
Current	505 (13.5)	561 (15.0)	173 (9.0)	263 (13.6)	213 (14.2)	280 (18.7)
Former	1609 (43.0)	1805 (48.3)	799 (41.4)	901 (46.7)	624 (41.6)	689 (45.9)
Never	1625 (43.5)	1373 (36.7)	956 (49.6)	764 (39.6)	663 (44.2)	531 (35.4)
ASA[b] class, n (%)						
1-2	421 (11.3)	143 (3.8)	126 (6.5)	37 (1.9)	250 (16.7)	81 (5.4)
3-4	3102 (83.0)	2875 (76.9)	1722 (89.3)	1649 (85.5)	1132 (75.5)	1152 (76.8)
5 or E	216 (5.8)	721 (19.3)	80 (4.1)	242 (12.6)	118 (7.9)	267 (17.8)
Surgical specialty, n (%)						
Cardiothoracic (CT)	536 (14.3)	577 (15.4)	183 (9.5)	160 (8.3)	72 (4.8)	142 (9.5)
Ears, nose, and throat (ENT)	48 (1.3)	80 (2.1)	76 (3.9)	98 (5.1)	17 (1.1)	77 (5.1)
General	498 (13.3)	490 (13.1)	952 (49.4)	981 (50.9)	309 (20.6)	370 (24.7)
Multiple	97 (2.6)	614 (16.4)	78 (4.0)	322 (16.7)	15 (1.0)	74 (4.9)
Neurology	666 (17.8)	672 (18.0)	3 (0.2)	10 (0.5)	169 (11.3)	128 (8.5)
Orthopedics	907 (24.3)	620 (16.6)	103 (5.3)	68 (3.5)	560 (37.3)	370 (24.7)
Other	28 (0.7)	28 (0.7)	57 (3.0)	61 (3.2)	11 (0.7)	22 (1.5)
Plastic surgery	165 (4.4)	111 (3.0)	31 (1.6)	17 (0.9)	77 (5.1)	95 (6.3)
Urology/gynecology	276 (7.4)	172 (4.6)	440 (22.8)	209 (10.8)	153 (10.2)	131 (8.7)
Vascular	518 (13.9)	375 (10.0)	5 (0.3)	2 (0.1)	117 (7.8)	91 (6.1)

[a]Continuous variables are summarized as the median (IQR) and categorical variables as n (%).

[b]ASA: American Society of Anesthesiologists.

Table 2. Clinical characteristics of cases and controls by institution.

Variable[a]	Institution A		Institution B		Institution C	
	Controls (n=3739)	Cases (n=3739)	Controls (n=1928)	Cases (n=1928)	Controls (n=1500)	Cases (n=1500)
ECI[b] score, median (IQR)	5 (0-13)	8 (2-18)	9 (4-17)	13 (5-22)	5 (0-12)	9 (2-18)
Number of ICD[c] codes, median (IQR)	21 (12-33)	24 (12-40)	21 (11-34)	26 (13-41)	17 (80-29)	22 (11-38)
Congestive heart failure (CHF), n (%)	713 (19.1)	1040 (27.8)	203 (10.5)	304 (15.8)	267 (17.8)	445 (29.7)
Arrhythmia, n (%)	969 (25.9)	1203 (32.2)	397 (20.6)	485 (25.2)	393 (26.2)	471 (31.4)
Valvular disease, n (%)	639 (17.1)	724 (19.4)	148 (7.7)	188 (9.8)	115 (7.7)	178 (11.9)
Peripheral vascular disorder (PVD), n (%)	977 (26.1)	1138 (30.4)	217 (11.3)	259 (13.4)	316 (21.1)	378 (25.2)
Hypertension, n (%)	2767 (74.0)	2696 (72.1)	1217 (63.1)	1255 (65.1)	997 (66.5)	1066 (71.1)
Chronic obstructive pulmonary disorder (COPD), n (%)	962 (25.7)	1227 (32.8)	444 (23.0)	506 (26.2)	398 (26.5)	542 (36.1)
Diabetes, n (%)	1295 (34.6)	1502 (40.2)	558 (28.9)	717 (37.2)	481 (32.1)	636 (42.4)
Hypothyroidism, n (%)	659 (17.6)	630 (16.8)	347 (18.0)	339 (17.6)	249 (16.6)	310 (20.7)
Renal failure, n (%)	891 (23.8)	1198 (32.0)	506 (26.2)	639 (33.1)	336 (22.4)	474 (31.6)
Liver disease, n (%)	266 (7.1)	336 (9.0)	415 (21.5)	573 (29.7)	73 (4.9)	129 (8.6)
Lymphoma, n (%)	69 (1.8)	85 (2.3)	75 (3.9)	71 (3.7)	33 (2.2)	26 (1.7)
Cancer, n (%)	986 (26.4)	1040 (27.8)	1273 (66.0)	1184 (61.4)	339 (22.6)	400 (26.7)
Coagulopathy, n (%)	264 (7.1)	393 (10.5)	164 (8.5)	336 (17.4)	100 (6.7)	153 (10.2)
Obesity, n (%)	720 (19.3)	758 (20.3)	303 (15.7)	372 (19.3)	333 (22.2)	374 (24.9)
Weight loss, n (%)	240 (6.4)	371 (9.9)	220 (11.4)	349 (18.1)	76 (5.1)	173 (11.5)
Fluid/electrolyte disorders, n (%)	761 (20.4)	1171 (31.3)	440 (22.8)	716 (37.1)	334 (22.3)	543 (36.2)
Deficiency anemia, n (%)	460 (12.3)	659 (17.6)	244 (12.7)	335 (17.4)	211 (14.1)	296 (19.7)
Alcohol abuse, n (%)	135 (3.6)	219 (5.9)	67 (3.5)	129 (6.7)	30 (2.0)	74 (4.9)
Drug abuse, n (%)	171 (4.6)	213 (5.7)	58 (3.0)	80 (4.1)	42 (2.8)	72 (4.8)
Psychoses, n (%)	20 (0.5)	84 (2.2)	13 (0.7)	34 (1.8)	12 (0.8)	38 (2.5)
Depression, n (%)	905 (24.2)	1022 (27.3)	343 (17.8)	514 (26.7)	275 (18.3)	406 (27.1)
CVD[d], n (%)	527 (14.1)	668 (17.9)	111 (5.8)	141 (7.3)	142 (9.5)	231 (15.4)
Previous TBI[e], n (%)	35 (0.9)	74 (2.0)	12 (0.6)	19 (1.0)	17 (1.1)	23 (1.5)
Sensory impairment, n (%)	212 (5.7)	203 (5.4)	81 (4.2)	91 (4.7)	75 (5.0)	118 (7.9)
Previous delirium, n (%)	215 (5.8)	615 (16.4)	103 (5.3)	304 (15.8)	85 (5.7)	278 (18.5)

[a]Continuous variables are summarized as the median (IQR) and categorical variables as n (%).

[b]ECI: Elixhauser comorbidity index.

[c]ICD: *International Classification of Diseases.*

[d]CVD: cerebrovascular disease.

[e]TBI: traumatic brain injury.

Model Evaluation

XGB had the highest AUROC out of the 4 candidate classifiers (AUROC=0.79), followed by the neural network (AUROC=0.78), the random forest (AUROC=0.78), and logistic regression (AUROC=0.72). Based on this AUROC evaluation, the XGB model was retained for further analysis. For institution A, the training set included 5234 visits (n=2617, 50%, cases and n=2617, 50%, controls) and the holdout set included 1503 visits (n=752, 50%, cases and n=751, 50%, controls). For institution B, the training and holdout data sets included 2699 visits (n=1350, 50%, cases and n=1349, 50%, controls) and 775 visits (n=387, 49.9%, cases and n=388, 50.1%, controls), respectively. The training and holdout data sets for institution C included 2100 visits (n=1050, 50%, cases and n=1050, 50%, controls) and 603 visits (n=302, 50.1%, cases and n=301, 49.9%, controls), respectively.

The models trained on institution A (ie, XGB_A) had the best performance, achieving AUROCs of 0.77-0.79 on institution A holdout data and 0.68-0.74 when externally validated on data from institutions B and C. Models trained on institution B (ie, XGB_B) were the least robust, achieving a maximum AUROC of 0.71 on holdout data from institution B and 0.72-0.74 when externally validated on data from institutions A and C. Models trained on institution C (ie, XGB_C) performed better than XGB_B but worse than XGB_A, with a maximum AUROC of 0.77 on holdout data from institution C and 0.64-0.75 when externally validated on data from institutions A and B (Table 3).

Table 3. XGB[a] model performance metrics[b] by surveillance period and holdout data.

Surveillance period, models, and institutions	AUROC[c], mean (SD)	Sensitivity, mean (SD)	Specificity, mean (SD)	PPV[d], mean (SD)	NPV[e], mean (SD)
1 year, XGB_A					
Institution A	0.79 (0.01)	0.70 (0.02)	0.72 (0.02)	0.72 (0.02)	0.71 (0.02)
Institution B	0.69 (0.02)	0.49 (0.03)	0.78 (0.02)	0.69 (0.03)	0.61 (0.02)
Institution C	0.74 (0.02)	0.70 (0.03)	0.66 (0.03)	0.67 (0.03)	0.69 (0.03)
1 year, XGB_B					
Institution A	0.74 (0.01)	0.76 (0.02)	0.57 (0.02)	0.64 (0.02)	0.70 (0.02)
Institution B	0.71 (0.02)	0.57 (0.03)	0.75 (0.02)	0.69 (0.03)	0.64 (0.02)
Institution C	0.73 (0.02)	0.65 (0.03)	0.68 (0.03)	0.67 (0.03)	0.66 (0.03)
1 year, XGC_C					
Institution A	0.75 (0.01)	0.75 (0.02)	0.60 (0.02)	0.66 (0.02)	0.71 (0.02)
Institution B	0.69 (0.02)	0.47 (0.03)	0.77 (0.02)	0.67 (0.03)	0.59 (0.02)
Institution C	0.77 (0.02)	0.66 (0.03)	0.69 (0.03)	0.69 (0.03)	0.67 (0.03)
6 months, XGB_A					
Institution A	0.78 (0.01)	0.56 (0.03)	0.73 (0.02)	0.67 (0.03)	0.62 (0.02)
Institution B	0.68 (0.02)	0.45 (0.03)	0.79 (0.02)	0.68 (0.03)	0.59 (0.02)
Institution C	0.74 (0.02)	0.67 (0.03)	0.66 (0.03)	0.67 (0.03)	0.67 (0.03)
6 months, XGB_B					
Institution A	0.73 (0.01)	0.78 (0.02)	0.54 (0.02)	0.63 (0.02)	0.71 (0.02)
Institution B	0.71 (0.02)	0.56 (0.03)	0.73 (0.02)	0.67 (0.03)	0.62 (0.02)
Institution C	0.74 (0.02)	0.66 (0.03)	0.68 (0.03)	0.68 (0.03)	0.67 (0.03)
6 months, XGC_C					
Institution A	0.73 (0.01)	0.76 (0.02)	0.55 (0.02)	0.63 (0.02)	0.70 (0.02)
Institution B	0.65 (0.02)	0.52 (0.03)	0.70 (0.02)	0.64 (0.03)	0.60 (0.02)
Institution C	0.76 (0.02)	0.71 (0.03)	0.66 (0.03)	0.68 (0.03)	0.69 (0.03)
3 months, XGB_A					
Institution A	0.77 (0.01)	0.70 (0.02)	0.70 (0.02)	0.70 (0.02)	0.70 (0.02)
Institution B	0.69 (0.02)	0.47 (0.03)	0.78 (0.02)	0.68 (0.03)	0.60 (0.02)
Institution C	0.74 (0.02)	0.68 (0.03)	0.67 (0.03)	0.67 (0.03)	0.68 (0.03)
3 months, XGB_B					
Institution A	0.72 (0.01)	0.75 (0.02)	0.55 (0.02)	0.63 (0.02)	0.69 (0.02)
Institution B	0.70 (0.02)	0.56 (0.03)	0.74 (0.02)	0.68 (0.03)	0.63 (0.02)
Institution C	0.74 (0.02)	0.65 (0.03)	0.68 (0.03)	0.67 (0.03)	0.66(0.03)
3 months, XGC_C					
Institution A	0.73 (0.01)	0.75 (0.02)	0.57 (0.02)	0.64 (0.02)	0.70 (0.02)
Institution B	0.64 (0.02)	0.50 (0.03)	0.71 (0.02)	0.63 (0.03)	0.58 (0.02)
Institution C	0.76 (0.02)	0.73 (0.03)	0.64 (0.03)	0.67 (0.03)	0.70 (0.03)

[a]XGB: extreme gradient boosting.

[b]Mean (SD) metrics presented were obtained using bootstrap resampling on the held-out patients from institutions A, B, and C.

[c]AUROC: area under the receiver operating curve.

[d]PPV: positive predictive value.

[e]NPV: negative predictive value.

Performance became marginally worse with shorter surveillance periods. All models were relatively well calibrated (Figures S1-S3 in Multimedia Appendix 1). The top 5 most important features for XGB_A, XGB_B, and XGB_C by evaluation data set and surveillance period are presented in Table 4 and Tables S8-S9 in Multimedia Appendix 1. The ASA class was frequently the most important predictor.

Across all surveillance periods, the models predicted between 40% and 60% of the patients with confusion as cases or controls (Table S10 in Multimedia Appendix 1).

Table 4. Top 5 most influential variables used by XGB[a] models (1-year surveillance period).[b]

Model and rank	Holdout data		
	Institution A	Institution B	Institution C
XGB_A			
1	ASA[c] class	ASA class	ASA class
2	ICD[d] group: Z00-Z13[e]	ICD group: Z00-Z13	ICD group: Z00-Z13
3	Multispecialty surgery	Multispecialty surgery	Service: hospitalist[f]
4	Service: hospitalist	Service: hospitalist	Multispecialty surgery
5	Emergency surgery	Previous delirium	Emergency surgery
XGB_B			
1	ASA class	ASA class	ASA class
2	Multispecialty surgery	Multispecialty surgery	Multispecialty surgery
3	Previous delirium	Previous delirium	Previous delirium
4	BMI	Urology/gynecology surgery	Service: orthopedics[g]
5	Emergency surgery	BMI	BMI
XGB_C			
1	ASA class	ASA class	ASA class
2	Service: hospitalist	Service: hospitalist	Service: orthopedics
3	Service: orthopedics	Service: orthopedics	Service: hospitalist
4	Previous delirium	Previous delirium	Previous delirium
5	Multispecialty surgery	Multispecialty surgery	ICD group: Z77-Z99[h]

[a]XGB: extreme gradient boosting.

[b]Feature importance measured using Shapley Additive Explanation (SHAP) values. XGB_A, XGB_B, and XGB_C were trained on data from institutions A, B, and C, respectively.

[c]ASA: American Society of Anesthesiologists.

[d]ICD: *International Classification of Diseases.*

[e]ICD group Z00-Z13: persons encountering health services for examinations.

[f]Admitted to hospitalist service.

[g]Admitted to orthopedics service.

[h]ICD group Z77-Z99: persons with potential health hazards related to family and personal history and certain conditions influencing health status.

Discussion

Principal Findings

We developed and externally validated 3 models to predict POD with routine EHR data available at the time of hospital admission. In our experiments, XGB outperformed all other classifiers and demonstrated good discriminative ability on holdout data, achieving a maximum AUROC of 0.79. Generalizability varied by model and the institution used for external validation.

Our models demonstrated good predictive accuracy, with XGB_A outperforming XGB_B and XGB_C across all surveillance periods. Interestingly, longer surveillance periods did not appear to significantly benefit model performance. This is likely because the most important features were surgery-related variables, which were fixed across all surveillance durations. Additionally, surveillance duration did not impact how the models classified patients with confusion but no delirium (ie, potential subsyndromal delirium); approximately half were predicted to be cases, and the other half were predicted to be controls, regardless of the surveillance period. Given that subsyndromal delirium is thought to be on the spectrum between healthy controls and delirium [38], it was expected that the models would have trouble classifying those patients.

Generalizability varied by model and institution. XGB_A performed relatively well when externally validated using data from institution C, as did XGB_C when validated using data from institution A. However, the AUROCs for both models decreased substantially when validated on data from institution B. In contrast, XGB_B had higher AUROCs when externally validated on institutions A and C than it did on holdout data from the same institution it was trained on. We hypothesize that the observed variation in performance could be due to institution B having a substantially different patient population than institutions A and C. Institutions A and C are trauma centers that perform a comparatively large number of orthopedic surgeries, and their populations have fewer comorbidities. Institution A also cares for complex vascular and cardiac patients, while the other 2 institutions generally do not. Conversely, institution B is not a trauma center and performs mostly general and urologic/gynecologic surgeries. It also largely services frail, high-acuity patients with chronic illnesses, and the general surgical complexity is higher. The comparatively low AUROC of XGB_B could reflect the model having difficulty discriminating between cases and controls, because it was trained on patients who were more ill, regardless of delirium status. These results highlight the importance of selecting an appropriate training population when a generalizable prediction model is desired; if a hospital has a patient population that differs significantly from the training data set, a localized model may be needed, even within the same hospital system.

The ASA class, a subjective measure of a patient's physiologic status [39], was frequently the most important feature. This supports previous literature linking a higher ASA class to a greater risk of POD [40]. The Elixhauser comorbidity index (ECI) did not appear in the list of top features despite the strong association of comorbidities with delirium, possibly because the ASA class summarizes health information beyond mortality risk and additionally identifies emergency cases. However, the subjectivity of the ASA class [41] may harm model generalizability compared to more objective measures, such as comorbidity scores. Other surgical variables, including admitting service and surgical specialty, were frequently among the top 5 features. Notably, both these variables have been associated with an increased risk of POD, particularly surgical specialty [6]. Multispecialty surgery was particularly important across models, suggesting that surgical complexity may be an important risk factor for delirium. The type of admitting service and individual surgical specialties that were most predictive differed by model, potentially because the distributions were different between institutions. For example, urologic/gynecologic surgery was frequently a top predictor in XGB_B models but not in others. This could be because proportionally more controls had that type of surgery than cases at institution B but not at institutions A and C. Reducing the cardinality of these variables is likely to improve generalizability but potentially at the cost of reduced discriminative ability. For XGB_A, the number of ICD codes belonging to ICD-10 group Z00-Z13 ("persons encountering health services for examinations") was a top feature, and higher values negatively influenced model predictions. This may be because this ICD group captures routine health examinations, which are often undertaken by healthier individuals. The fact that the top features are supported by the literature suggests that the models are clinically explainable.

Several delirium prediction models have been developed, reporting AUROCs ranging from 0.56 to 0.94 [42]. The models with the highest AUROCs focus on specific patient subsets (ie, ICU patients, cardiac surgery) and include variables collected during the hospital stay, such as the APACHE score (which must be calculated), surgery duration (often not reliably recorded), and inpatient laboratory values. In-hospital variables may, indeed, be the strongest predictors of delirium and explain why our model failed to outperform previous ones; however, they were intentionally excluded from this study as that would preclude our models from being used at the time of hospitalization. Fewer models have been developed that are both her based and intended to be used at or shortly after admission. In their 2022 paper, Bishara et al [14] developed a POD prediction model for the general surgical population using different machine learning approaches and preoperative EHR data. They found that an XGB model outperforms other classifiers, similar to our findings, and reported an internal validation AUROC of 0.85 [14]. In contrast to our study, matching was not performed, and patients with dementia were included in the study population. Fifty-nine variables derived from inpatient (but preoperative) nursing assessments were also included as predictors. Some of these assessments (eg, Braden Scale score [43]) captured patients' functional status, which is highly correlated to delirium [5,6] and may explain why their model had a higher AUROC. Wong et al [44] developed a model to predict delirium in a general inpatient population without known cognitive impairment using an XGB model and reported an AUROC of 0.86. Their model used 796 features collected within 24 hours of admission and included inpatient neurologic examination data, which were highly predictive of delirium. These factors could explain, at least in part, the difference in performance between these previous models and our models.

In summary, our findings suggest that a machine learning model trained on routine EHR data can achieve clinically useful accuracy when predicting POD. Unlike previous models, the models presented in this study can be used to make predictions at the time of hospital admission, which could quickly inform preventive and resource-planning efforts. The models were also externally validated, providing critical information about generalizability when using a limited set of prehospital and surgery variables. These models can be readily integrated into EHR systems to provide a scalable, automated prescreening tool to flag patients who are at risk of developing POD and would benefit from targeted preventative measures.

Strengths and Limitations

Our study has several strengths. First, we used both the CAM method and ICD codes to maximize case identification; because delirium ICD codes are extremely specific but less sensitive [45], false negatives are unlikely. Second, we compared different surveillance periods to determine how surveillance duration influences accuracy. Third, we examined how the models classify patients with confusion but no delirium, which could potentially capture subsyndromal delirium. Finally, we trained our models on data from 3 different institutions and externally validated them against each other to determine their transportability.

This study also has several limitations. Although we attempted to maximize delirium detection by using both the CAM method and ICD codes, a small number of patients did not have any CAM data available. As mentioned previously, delirium ICD codes tend to have high specificity but lower sensitivity [45], so some cases may have been missed. Patients were intentionally matched on age, sex, and race to limit biases related to these variables; however, discriminative ability was likely reduced as a result. Because patients with preexisting dementia or confusion during the inpatient visit (but no documented delirium) were excluded, the models may not generalize well to those types of patients. However, we chose to exclude those patients because their high risk of delirium was evident; our models focused on patients with a less clear delirium risk, which could partially explain the lower performance compared to previous models. Finally, although the models were externally validated, the hospitals were within the same health care system, which may present more optimistic generalizability relative to uses of the models in outside systems.

Conclusion

Routine EHR data can be used for early delirium prediction in a diverse cohort of surgery patients without dementia. Although our models slightly underperformed relative to some of the previously published classifiers that use inpatient data, our routine EHR-based models serve a distinct purpose of enabling predictions at the time of admission, while being highly scalable. Generalizability varied depending on the training data, so institution-specific models may be necessary when using only a limited set of preadmission and surgery variables with distributions that substantially differ between institutions. The proposed models could be used in clinical practice as an automated prescreening tool for the early identification of high-risk patients, enabling clinicians to immediately adjust their care strategies and inform targeted delirium prevention measures and resource planning.

Acknowledgments

This study was supported by the National Institute on Aging (#K23AG071945).

Authors' Contributions

All authors contributed to study conception and design. EH performed all analyses and drafted the manuscript. All authors critically revised and reviewed the final manuscript.

Conflicts of Interest

ZBM has a financial interest in DigiCare Realized and could benefit from the results of this research. MB serves as a chief scientific officer and cofounder of BlueAgilis; and the chief health officer of DigiCare Realized, Inc. He has equity interest in Blue Agilis, Inc DigiCare Realized, Inc; Preferred Population Health Management LLC; and MyShift, Inc (previously known as RestUp, LLC). He serves as an advisory board member for Acadia Pharmaceuticals; Eisai, Inc; Biogen; and Genentech. These conflicts have been reviewed by Indiana University and have been appropriately managed to maintain objectivity. The remaining authors declare no competing interests.

Multimedia Appendix 1

Calibration curves for XGB; ICD codes for preexisting Alzheimer's disease, related dementias, delirium, and additional variables; sociodemographic and surgical characteristics of patients and controls; clinical characteristics of patients and controls; XGB model predictions for confusion encounters; and top 5 most influential variables used by XGB models. ICD: International Classification of Diseases; XGB: extreme gradient boosting.

[DOCX File , 271 KB - periop_v8i1e59422_app1.docx]

References

1. Marcantonio ER, Goldman L, Mangione CM, Ludwig LE, Muraca B, Haslauer CM, et al. A clinical prediction rule for delirium after elective noncardiac surgery. JAMA 1994 Jan 12;271(2):134-139. [Medline: 8264068]
2. Marcantonio ER, Flacker JM, Wright RJ, Resnick NM. Reducing delirium after hip fracture: a randomized trial. J Am Geriatr Soc 2001 May;49(5):516-522. [doi: 10.1046/j.1532-5415.2001.49108.x] [Medline: 11380742]
3. Rudolph JL, Jones RN, Levkoff SE, Rockett C, Inouye SK, Sellke FW, et al. Derivation and validation of a preoperative prediction rule for delirium after cardiac surgery. Circulation 2009 Jan 20;119(2):229-236 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.108.795260] [Medline: 19118253]
4. Inouye SK, Westendorp RGJ, Saczynski JS. Delirium in elderly people. Lancet 2014 Mar 08;383(9920):911-922 [FREE Full text] [doi: 10.1016/S0140-6736(13)60688-1] [Medline: 23992774]
5. Marcantonio ER. In the clinic. Delirium. Ann Intern Med 2011 Jun 07;154(11):ITC6-1, ITC6. [doi: 10.7326/0003-4819-154-11-201106070-01006] [Medline: 21646553]

6. Vasilevskis EE, Han JH, Hughes CG, Ely EW. Epidemiology and risk factors for delirium across hospital settings. Best Pract Res Clin Anaesthesiol 2012 Sep;26(3):277-287 [FREE Full text] [doi: 10.1016/j.bpa.2012.07.003] [Medline: 23040281]
7. Brouquet A, Cudennec T, Benoist S, Moulias S, Beauchet A, Penna C, et al. Impaired mobility, ASA status and administration of tramadol are risk factors for postoperative delirium in patients aged 75 years or more after major abdominal surgery. Ann Surg 2010 Apr;251(4):759-765. [doi: 10.1097/SLA.0b013e3181c1cfc9] [Medline: 20224380]
8. Greene NH, Attix DK, Weldon BC, Smith PJ, McDonagh DL, Monk TG. Measures of executive function and depression identify patients at risk for postoperative delirium. Anesthesiology 2009 Apr;110(4):788-795 [FREE Full text] [doi: 10.1097/aln.0b013e31819b5ba6] [Medline: 19326494]
9. Noimark D. Predicting the onset of delirium in the post-operative patient. Age Ageing 2009 Jul;38(4):368-373. [doi: 10.1093/ageing/afp024] [Medline: 19297372]
10. Koebrugge B, van Wensen RJA, Bosscha K, Dautzenberg PLJ, Koning OHJ. Delirium after emergency/elective open and endovascular aortoiliac surgery at a surgical ward with a high-standard delirium care protocol. Vascular 2010;18(5):279-287. [doi: 10.2310/6670.2010.00052] [Medline: 20822723]
11. Witlox J, Eurelings LSM, de Jonghe JFM, Kalisvaart KJ, Eikelenboom P, van Gool WA. Delirium in elderly patients and the risk of postdischarge mortality, institutionalization, and dementia: a meta-analysis. JAMA 2010 Jul 28;304(4):443-451. [doi: 10.1001/jama.2010.1013] [Medline: 20664045]
12. Inouye SK, Bogardus ST, Charpentier PA, Leo-Summers L, Acampora D, Holford TR, et al. A multicomponent intervention to prevent delirium in hospitalized older patients. N Engl J Med 1999 Mar 04;340(9):669-676. [doi: 10.1056/NEJM199903043400901] [Medline: 10053175]
13. Hshieh TT, Yang T, Gartaganis SL, Yue J, Inouye SK. Hospital elder life program: systematic review and meta-analysis of effectiveness. Am J Geriatr Psychiatry 2018 Oct;26(10):1015-1033 [FREE Full text] [doi: 10.1016/j.jagp.2018.06.007] [Medline: 30076080]
14. Bishara A, Chiu C, Whitlock EL, Douglas VC, Lee S, Butte AJ, et al. Postoperative delirium prediction using machine learning models and preoperative electronic health record data. BMC Anesthesiol 2022 Jan 03;22(1):8 [FREE Full text] [doi: 10.1186/s12871-021-01543-y] [Medline: 34979919]
15. Hu X, Liu H, Zhao X, Sun X, Zhou J, Gao X, et al. Automated machine learning-based model predicts postoperative delirium using readily extractable perioperative collected electronic data. CNS Neurosci Ther 2022 Apr;28(4):608-618 [FREE Full text] [doi: 10.1111/cns.13758] [Medline: 34792857]
16. Davoudi A, Ebadi A, Rashidi P, Ozrazgat-Baslanti T, Bihorac A, Bursian AC. Delirium prediction using machine learning models on preoperative electronic health records data. Proc IEEE Int Symp Bioinformatics Bioeng 2017 Oct;2017:568-573 [FREE Full text] [doi: 10.1109/BIBE.2017.00014] [Medline: 30393788]
17. Racine AM, Tommet D, D'Aquila ML, Fong TG, Gou Y, Tabloski PA, RISE Study Group. Machine learning to develop and internally validate a predictive model for post-operative delirium in a prospective, observational clinical cohort study of older surgical patients. J Gen Intern Med 2021 Feb;36(2):265-273 [FREE Full text] [doi: 10.1007/s11606-020-06238-7] [Medline: 33078300]
18. Xue B, Li D, Lu C, King CR, Wildes T, Avidan MS, et al. Use of machine learning to develop and evaluate models using preoperative and intraoperative data to identify risks of postoperative complications. JAMA Netw Open 2021 Mar 01;4(3):e212240 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.2240] [Medline: 33783520]
19. Zhao H, You J, Peng Y, Feng Y. Machine learning algorithm using electronic chart-derived data to predict delirium after elderly hip fracture surgeries: a retrospective case-control study. Front Surg 2021;8:634629 [FREE Full text] [doi: 10.3389/fsurg.2021.634629] [Medline: 34327210]
20. Matsumoto K, Nohara Y, Sakaguchi M, Takayama Y, Fukushige S, Soejima H, et al. Temporal generalizability of machine learning models for predicting postoperative delirium using electronic health record data: model development and validation study. JMIR Perioper Med 2023 Oct 26;6:e50895 [FREE Full text] [doi: 10.2196/50895] [Medline: 37883164]
21. Mufti HN, Hirsch GM, Abidi SR, Abidi SSR. Exploiting machine learning algorithms and methods for the prediction of agitated delirium after cardiac surgery: models development and validation study. JMIR Med Inform 2019 Oct 23;7(4):e14993 [FREE Full text] [doi: 10.2196/14993] [Medline: 31558433]
22. Jung JW, Hwang S, Ko S, Jo C, Park HY, Han H, et al. A machine-learning model to predict postoperative delirium following knee arthroplasty using electronic health records. BMC Psychiatry 2022 Jun 27;22(1):436 [FREE Full text] [doi: 10.1186/s12888-022-04067-y] [Medline: 35761274]
23. Luo Y, Wu X, Song Y, Wang X, Liu K, Shi C, et al. Development and validation of a nomogram to predict postoperative delirium in older patients after major abdominal surgery: a retrospective case-control study. Perioper Med (Lond) 2024 May 16;13(1):41 [FREE Full text] [doi: 10.1186/s13741-024-00399-3] [Medline: 38755693]
24. Geßele C, Saller T, Smolka V, Dimitriadis K, Amann U, Strobach D. Development and validation of a new drug-focused predictive risk score for postoperative delirium in orthopaedic and trauma surgery patients. BMC Geriatr 2024 May 13;24(1):422 [FREE Full text] [doi: 10.1186/s12877-024-05005-1] [Medline: 38741037]
25. Fan Y, Yang T, Liu Y, Gan H, Li X, Luo Y, et al. Nomogram for predicting the risk of postoperative delirium in elderly patients undergoing orthopedic surgery. Perioper Med (Lond) 2024 May 04;13(1):34 [FREE Full text] [doi: 10.1186/s13741-024-00393-9] [Medline: 38702728]

26. Nagata C, Hata M, Miyazaki Y, Masuda H, Wada T, Kimura T, et al. Development of postoperative delirium prediction models in patients undergoing cardiovascular surgery using machine learning algorithms. Sci Rep 2023 Nov 30;13(1):21090 [FREE Full text] [doi: 10.1038/s41598-023-48418-5] [Medline: 38036664]
27. Collins GS, Reitsma JB, Altman DG, Moons K. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMC Med 2015 Jan 06;13(1):1 [FREE Full text] [doi: 10.1186/s12916-014-0241-z] [Medline: 25563062]
28. IU Health: about our system. Indiana University Health. URL: https://iuhealth.org/about-our-system [accessed 2024-04-15]
29. Inouye SK, van Dyck CH, Alessi CA, Balkin S, Siegal AP, Horwitz RI. Clarifying confusion: the Confusion Assessment Method. A new method for detection of delirium. Ann Intern Med 1990 Dec 15;113(12):941-948. [doi: 10.7326/0003-4819-113-12-941] [Medline: 2240918]
30. Wei LA, Fearing MA, Sternberg EJ, Inouye SK. The Confusion Assessment Method: a systematic review of current usage. J Am Geriatr Soc 2008 May;56(5):823-830 [FREE Full text] [doi: 10.1111/j.1532-5415.2008.01674.x] [Medline: 18384586]
31. Khuri SF, Henderson WG, Daley J, Jonasson O, Jones RS, Campbell DA, Principal Investigators of the Patient Safety in Surgery Study. Successful implementation of the Department of Veterans Affairs' National Surgical Quality Improvement Program in the private sector: the Patient Safety in Surgery study. Ann Surg 2008 Aug;248(2):329-336. [doi: 10.1097/SLA.0b013e3181823485] [Medline: 18650645]
32. Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi J, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. Med Care 2005 Nov;43(11):1130-1139. [doi: 10.1097/01.mlr.0000182534.19832.83] [Medline: 16224307]
33. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. Med Care 1998 Jan;36(1):8-27. [doi: 10.1097/00005650-199801000-00004] [Medline: 9431328]
34. van Walraven C, Austin PC, Jennings A, Quan H, Forster AJ. A modification of the Elixhauser comorbidity measures into a point system for hospital death using administrative data. Med Care 2009 Jun;47(6):626-633. [doi: 10.1097/MLR.0b013e31819432e5] [Medline: 19433995]
35. Campbell N, Maidment I, Fox C, Khan B, Boustani M. The 2012 update to the anticholinergic cognitive burden scale. J Am Geriatr Soc 2013 Apr;61(S1):S142-S143. [doi: 10.1111/JGS.2013.61.ISSUE-S1]
36. Anatomical therapeutic chemical (ATC) classification. World Health Organization. URL: https://www.who.int/tools/atc-ddd-toolkit/atc-classification [accessed 2023-09-04]
37. Lundberg SM, Lee LSI. A unified approach to interpreting model predictions. arXiv. Preprint posted online 2012. [doi: 10.48550/arXiv.1705.07874] . [doi: 10.48550/arXiv.1705.07874]
38. Ouimet S, Riker R, Bergeron N, Cossette M, Kavanagh B, Skrobik Y. Subsyndromal delirium in the ICU: evidence for a disease spectrum. Intensive Care Med 2007 Jun 3;33(6):1007-1013. [doi: 10.1007/s00134-007-0618-y] [Medline: 17404704]
39. Doyle DJ, Hendrix JM, Garmon EH. American Society of Anesthesiologists Classification. Treasure Island, FL: StatPearls Publishing; 2023.
40. Oh E, Li M, Fafowora T, Inouye S, Chen C, Rosman L, et al. Preoperative risk factors for postoperative delirium following hip fracture repair: a systematic review. Int J Geriatr Psychiatry 2015 Sep;30(9):900-910 [FREE Full text] [doi: 10.1002/gps.4233] [Medline: 25503071]
41. Sankar A, Johnson SR, Beattie WS, Tait G, Wijeysundera DN. Reliability of the American Society of Anesthesiologists physical status scale in clinical practice. Br J Anaesth 2014 Sep;113(3):424-432 [FREE Full text] [doi: 10.1093/bja/aeu100] [Medline: 24727705]
42. Lindroth H, Bratzke L, Purvis S, Brown R, Coburn M, Mrkobrada M, et al. Systematic review of prediction models for delirium in the older adult inpatient. BMJ Open 2018 Apr 28;8(4):e019223 [FREE Full text] [doi: 10.1136/bmjopen-2017-019223] [Medline: 29705752]
43. Bergstrom N, Braden BJ, Laguzza A, Holman V. The Braden Scale for predicting pressure sore risk. Nurs Res 1987;36(4):205-210. [Medline: 3299278]
44. Wong A, Young AT, Liang AS, Gonzales R, Douglas VC, Hadley D. Development and validation of an electronic health record-based machine learning model to estimate delirium risk in newly hospitalized patients without known cognitive impairment. JAMA Netw Open 2018 Aug 03;1(4):e181018 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.1018] [Medline: 30646095]
45. Kim D, Lee J, Kim C, Huybrechts K, Bateman B, Patorno E, et al. Evaluation of algorithms to identify delirium in administrative claims and drug utilization database. Pharmacoepidemiol Drug Saf 2017 Aug;26(8):945-953 [FREE Full text] [doi: 10.1002/pds.4226] [Medline: 28485014]

Abbreviations

ACB: Anticholinergic Cognitive Burden
ACh: anticholinergic
ASA: American Society of Anesthesiologists
ATC: Anatomical Therapeutic Chemical

AUROC: area under the receiver operating characteristic curve
CAM: Confusion Assessment Method
CVD: cerebrovascular disease
ECI: Elixhauser comorbidity index
EHR: electronic health record
ICD: International Classification of Diseases
ICD-9: International Classification of Diseases, Ninth Revision
ICD-10-CM: International Classification of Diseases, Tenth Revision, Clinical Modification
IU: Indiana University
NPV: negative predictive value
POD: postoperative delirium
PPV: positive predictive value
SHAP: Shapley Additive Explanation
TBI: traumatic brain injury
XGB: extreme gradient boosting

Edited by N Rohatgi; submitted 11.04.24; peer-reviewed by SR Pagali; comments to author 26.07.24; revised version received 15.10.24; accepted 01.11.24; published 09.01.25.

Please cite as:
Holler E, Ludema C, Ben Miled Z, Rosenberg M, Kalbaugh C, Boustani M, Mohanty S
Development and Validation of a Routine Electronic Health Record-Based Delirium Prediction Model for Surgical Patients Without Dementia: Retrospective Case-Control Study
JMIR Perioper Med 2025;8:e59422
URL: https://periop.jmir.org/2025/1/e59422
doi:10.2196/59422
PMID:

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org