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

Review

Corrigenda and Addendas

Original Papers

Clinical Pathways in Knee and Hip Arthroplasty: Narrative Review on Sustainability, Quality, and Resource Management

Manuel Godinho[1], MD; Filipe Maçães[2], MD; Helena Gonçalves[3], MSc; Firmino Silva[3,4], PhD

[1]Departamento de Ortopedia e Traumatologia, Unidade Local de Saúde de Entre o Douro e Vouga, Rua. Dr. Cândido Pinho 5, Santa Maria da Feira, Portugal

[2]Departamento de Ortopedia e Traumatologia, Unidade Local de Saúde de Gaia e Espinho, Vila Nova de Gaia, Portugal

[3]ISLA - Polytechnic Institute of Management and Technology, Research Center, Vila Nova de Gaia, Portugal

[4]COPELABS – Cognitive and People-centric Computing Research Unit, Lisbon, Portugal

Corresponding Author:
Manuel Godinho, MD
Departamento de Ortopedia e Traumatologia, Unidade Local de Saúde de Entre o Douro e Vouga, Rua. Dr. Cândido Pinho 5, Santa Maria da Feira, Portugal

Abstract

Background: Increasing arthroplasty volumes are testing health care system capacity, budgets, and workforce resilience. Clinical pathways (CPWs) provide a practical, evidence-based structure that aligns perioperative actions from preparation through follow-up. In this review, we treat three aims as coprimary: quality (patient outcomes and adherence to best practice); resource management and efficiency at the episode level (eg, length of stay, perioperative flow, direct costs); and sustainability, defined as the ability to maintain high-quality services over time by optimizing financial, human, and environmental resources while safeguarding equitable access.

Objective: This study aimed to describe the main CPW subtypes used in hip and knee arthroplasty and synthesize evidence on their effects on quality of care, resource management, and sustainability.

Methods: We conducted a narrative review of studies indexed in PubMed and Cochrane (2013 - 2024) that evaluated CPWs in total hip and knee arthroplasty. Screening and selection were documented with a PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses)-style flow diagram for transparency, and findings were synthesized thematically.

Results: Across CPW models, consistent signals of benefit were identified. Enhanced Recovery After Surgery (ERAS) pathways accelerate recovery and enable earlier discharge without increasing complications, often reducing opioid exposure and time to mobilization. Integrated Clinical Pathways improve standardization and multidisciplinary coordination across settings, reducing unwarranted variability and supporting safer transitions of care. Fast-track programs emphasize early mobilization and streamlined perioperative processes, improving patient flow and satisfaction while decreasing length of stay. Outpatient arthroplasty pathways allow same-day discharge in carefully selected, low-risk patients, reducing bed occupancy and freeing inpatient capacity. Virtual clinics support remote follow-up, patient education, and complication surveillance, decreasing unnecessary in-person visits and optimizing clinician time. Collectively, these pathways align quality improvement with sustainability by lowering bed-days, improving adherence to evidence-based practices, and enabling more efficient use of operating rooms, wards, and workforce.

Conclusions: This review highlights the importance of CPWs in improving care delivery and patient outcomes in orthopedic surgery. Future efforts should refine CPWs, integrate digital tools and platforms, adopt standardized sustainability metrics, and stay flexible to evolving service demands.

(JMIR Perioper Med 2025;8:e78174) doi:10.2196/78174

KEYWORDS

clinical pathways; knee arthroplasty; hip arthroplasty; length of stay; treatment duration; treatment cost; treatment efficiency

Introduction

Background

Hip and knee arthroplasty volumes continue to rise worldwide, driven by population aging and lifestyle-related osteoarthritis. While these operations reliably restore function and relieve pain, their scale and cost intensity place sustained pressure on inpatient beds, operating theaters, budgets, and the clinical workforce, making the organization of care as critical as the technical act itself [1,2].

Clinical pathways (CPWs) provide an evidence-based, multidisciplinary structure that specifies what is done, when, and by whom across the perioperative continuum—from preparation and anesthesia and analgesia to mobilization,

discharge readiness, and follow-up. By reducing unwarranted variation and embedding best practice locally, CPWs have been associated in arthroplasty with shorter length of stay, stable short-term safety, better adherence to recommended processes, more reliable transitions of care, and more predictable use of resources across settings [3-8].

In this context, we frame three coprimary aims for CPWs. Quality addresses patient outcomes and fidelity to evidence-based practice. Resource management and efficiency captures near-term, episode- or service-line performance (eg, length of stay, perioperative flow, discharge disposition, direct costs). Sustainability reflects the system's ability to maintain high-quality services over time by optimizing financial, human, and environmental resources while safeguarding equitable access, conceptually distinct from, but complementary to, efficiency. As health systems adopt value-based models, CPWs offer a practical vehicle to advance all three aims within routine delivery, rather than trading one for another [9,10].

Patient-centered pathway variants further illustrate this alignment. Fast-track programs pair clear milestones and early mobilization with person-centered communication, supporting timely, safe discharge and high satisfaction; outpatient arthroplasty pathways extend this logic to same-day discharge in appropriately selected patients, relieving bed occupancy without an apparent safety penalty. Digital adjuncts—including protocolized virtual follow-up, structured remote review, and routine PROM capture, as well as patient infotainment systems for education—can standardize escalation criteria, triage unnecessary face-to-face visits, and help preserve in-person capacity for those who need it most [11-16].

Prior Work

Beaupre et al [17] describe clinical pathways (CPWs) as frameworks that promote adherence to best practice and turn guidelines into coordinated local care. By making clear who does what, when, and for whom, CPWs reduce unwarranted variation between providers and help maintain consistent care quality across the perioperative journey—from preparation to recovery and follow-up [18].

To maintain clarity in terms in this review, we group the literature into five pathway models that are widely used in orthopaedics: Enhanced Recovery After Surgery (ERAS) [19,20] (evidence-based perioperative bundles to attenuate surgical stress and speed recovery); Integrated Clinical Pathways (ICPs) [21] (structured coordination across teams and care settings); Fast-track pathways [22] (early mobilisation, optimised multimodal analgesia, and clear daily milestones); Outpatient arthroplasty pathways [23] (same-day discharge for appropriately selected patients); and Virtual clinics [13] (protocol-driven remote follow-up and patient education). Although these labels sometimes overlap in the literature, we use them consistently as distinct operational models in this review; brief definitions appear in Table 1.

ERAS pathways prioritize multidisciplinary, evidence-based perioperative care to reduce surgical stress, shorten hospital stays, and accelerate functional recovery, particularly in elective arthroplasty cases. ICPs aim to coordinate care across settings and providers, improving consistency and optimizing resource use in complex surgeries requiring long-term follow-up. Fast-track pathways focus on early mobilization, minimally invasive techniques, and optimized pain management, facilitating safe and rapid discharge for low-risk patients. Outpatient Arthroplasty pathways support same-day discharge in selected patients, helping manage surgical backlogs and reduce costs while maintaining safety and satisfaction. Virtual clinics offer remote follow-up, patient education, and digital assessments, enhancing convenience and reducing unnecessary in-person visits. Each model contributes to the broader goal of delivering efficient, standardized, and patient-centered care in joint arthroplasty while addressing different facets of the surgical journey.

Table . Concise summary of each CPW subtype.

CPW subtype	Main features	Primary goals	Typical applications	Example benefits*
ERAS (Enhanced Recovery After Surgery) [19,20]	Multidisciplinary, evidence-based perioperative protocols to speed recovery	Reduce hospital stay, accelerate recovery	Elective hip and knee arthroplasty	Discharge in 0 - 3 days, no increased complications [6,24]
ICP (Integrated Clinical Pathway) [21]	Structured care plans ensuring coordination across teams and settings	Improve consistency, reduce variability, optimize resources	Complex surgeries with extended follow-up	Reduced length of stay and costs, mixed results [7]
Fast-track Pathway [22[]	Protocols focusing on early mobilization and optimized pain control	Reduce hospital stay, reduce complications	Joint replacement, low-risk patients	Earlier discharge, fewer complications [22]
Outpatient Arthroplasty Pathway [23]	Protocols for same-day discharge in selected patients	Maximize efficiency, reduce costs	Elective arthroplasty, low-risk patients	85% same-day discharge, cost savings [12]
Virtual Clinic [13]	Digital platforms for remote follow-up and education	Improve convenience, reduce in-person visits	Post-operative follow-up in orthopedics	Fewer in-person visits, improved satisfaction [14]

Objectives

This narrative review examines how CPWs in hip and knee arthroplasty affect three vectors: quality of care, resource management (efficiency), and sustainability.

Quality refers to clinical and patient-reported outcomes and adherence to evidence-based practices.

Efficiency denotes near-term, episode- or service-line performance (eg, length of stay, operating-room turnover, discharge readiness, episode cost).

Sustainability is the health system's capacity to maintain high-quality arthroplasty care over time while optimizing financial, human, and environmental resources and preserving equitable access. To avoid conflation, efficiency is treated as a short-term operational outcome, whereas sustainability captures durable, system-level impact across economic, workforce, environmental (proxy), and equity domains. Figure 1 illustrates the three-domain framework (quality, efficiency, sustainability) that guides this review.

Figure 1. Three fundamental vectors in health care.

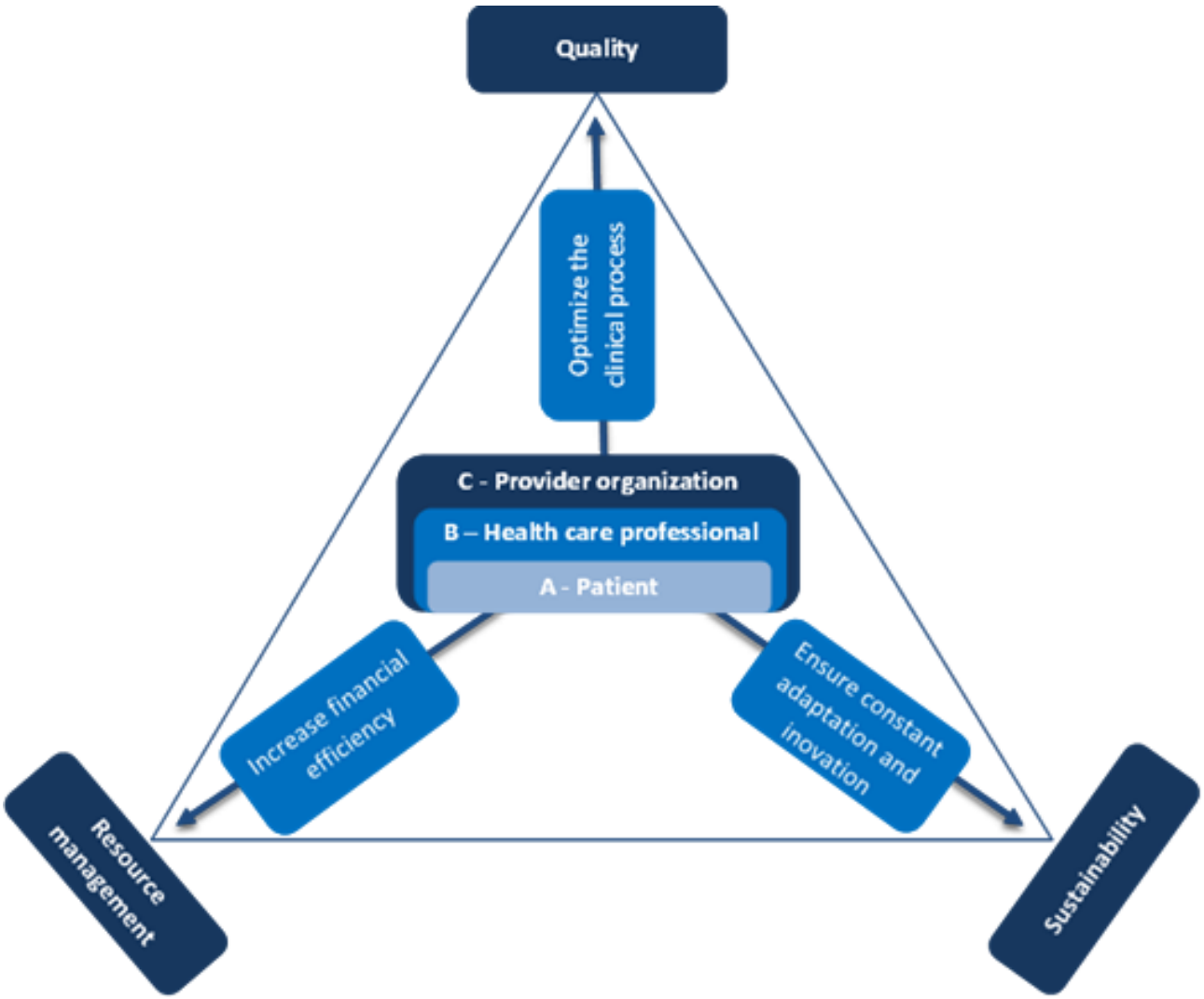


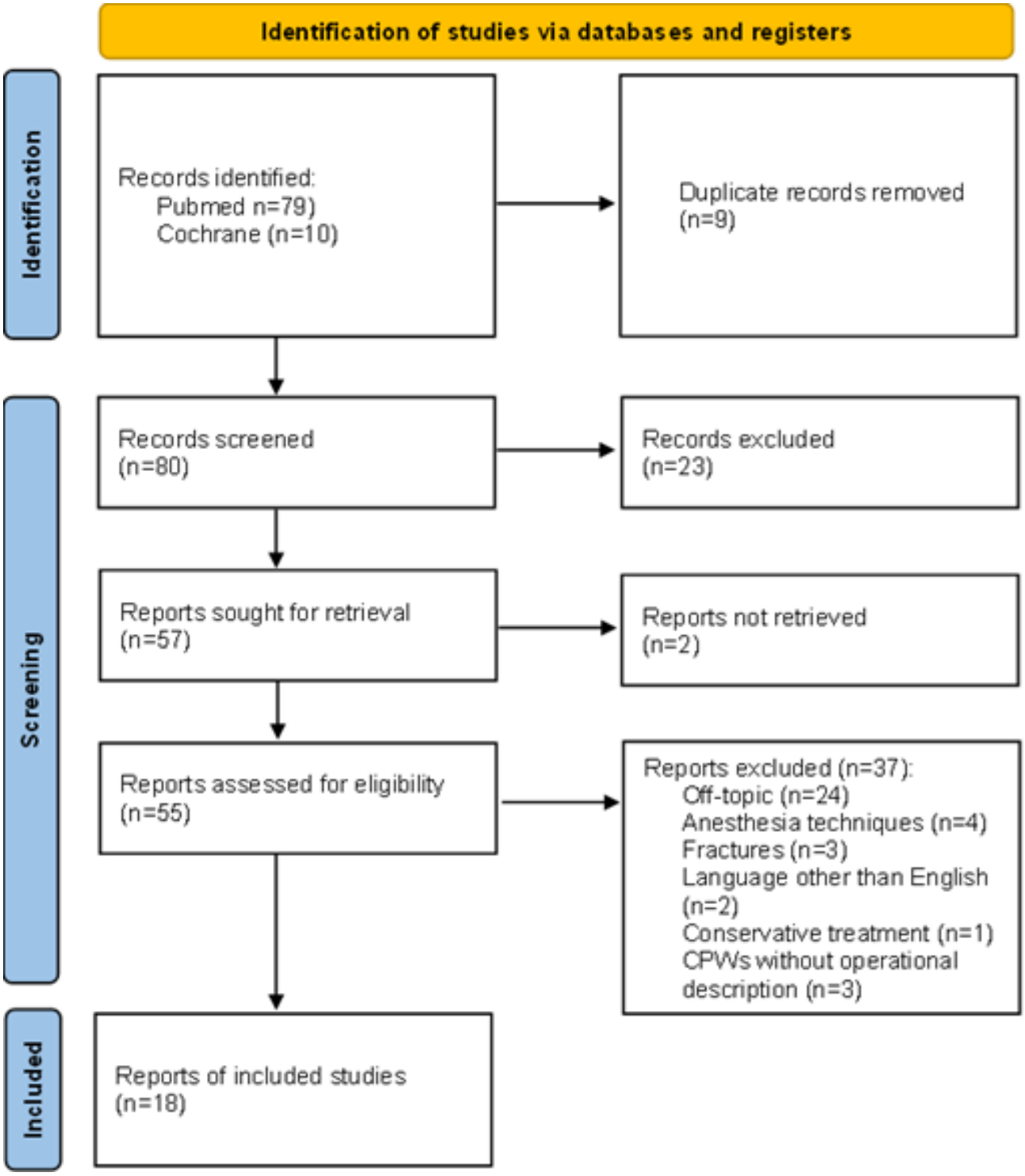
Figure 1 Conceptual framework: clinical pathways optimize the clinical process at three stakeholder levels: patient, health care professional, and provider organization to align service quality, resource management (efficiency), and health-system sustainability.

Methods

Literature Search

To explore the impact of CPWs in knee and hip arthroplasty, we conducted a literature search using PubMed and Cochrane databases, focusing on articles published in the past 12 years (January 2013 to December 2024). We used the MeSH terms "Critical Pathways"[Mesh] AND "Arthroplasty"[Mesh], including only English-language articles. While PRISMA [25] guidelines are typically used for systematic reviews, we adapted them to ensure a transparent and systematic literature search, given the diverse clinical pathways in joint arthroplasty. Studies were included if they addressed CPWs (including ERAS, ICPs, fast-track, outpatient arthroplasty pathways, or virtual clinics) for knee or hip arthroplasty, and reported outcomes related to quality, resource management or sustainability. Of the 89 articles identified, 18 met these criteria and were included in the review. The full study selection process and reasons for exclusion are detailed in Figure 2 (PRISMA flow diagram).

Figure 2. PRISMA flowchart diagram; CPWs - clinical pathways.



Inclusion Criteria

Studies of clinical pathways for hip arthroplasty, knee arthroplasty, or both, that reported outcomes related to sustainability, quality, or resource management were included.

Exclusion Criteria

Studies not addressing clinical pathways in hip or knee arthroplasty; articles not retrievable; articles not available in English were excluded.

Data Extraction and Synthesis

Each included study was classified into one of five predefined pathway models (Table 1): ERAS, ICPs, fast-track programs, outpatient arthroplasty pathways, or virtual clinics. Mixed reports were assigned to the predominant subtype based on operational characteristics. Owing to heterogeneity in designs and outcomes, we extracted data as reported and synthesized findings narratively, organizing results under three co-primary aims: quality, resource management (efficiency), and sustainability (mapped to economic, operational, workforce, environmental-proxy, and equity or access domains).

Ethical Considerations

This narrative review used only previously published data, with no new human subject involvement; therefore, Institutional Review Board review approval was not required.

Results

Overview of Clinical Pathway Subtypes

Across the 18 included studies, evidence addressed five predefined CPW models (Table 1): ERAS, ICPs, fast-track programs, outpatient arthroplasty pathways, and virtual clinics. Hybrid reports were assigned to the predominant subtype. Despite variation in scope and implementation, they share common goals of improving recovery, optimizing resource use, and standardizing care.

Quality of Care

Applying standardized practices through clinical pathways reduces variability and increases consistency, both critical for successful elective joint replacement. This is underscored by the "halo effect" observed with care pathways, where improvements in one procedure translate to gains in others [26].

Within enhanced recovery programs, quality improvements are seen in adherence to recommended practices (for example, prophylactic antibiotics within 60 minutes and early postoperative physiotherapy), which support better processes and outcomes. Foni et al [5] report gains, including reduced length of stay and stronger adherence to best practices, changes that align with improved patient experience and satisfaction.

ICPs elevate postoperative quality by standardizing protocols across the continuum of care, which likely contributes to fewer complications and shorter hospital stays, with corresponding improvements in perceived care quality. Zhou et al [7] highlight favorable effects on quality and efficiency metrics; although direct satisfaction measures were not detailed, the outcome profile implies enhanced patient perceptions. Complementing this, preoperative physical therapy, often embedded within standardized pathways, has been associated with a marked reduction in use of post-acute services; Snow et al [27] documented a 29% decrease across skilled nursing facilities, home health, and inpatient rehabilitation, reinforcing a recovery model that promotes patient independence.

Fast-track programs maintain high-quality care through a person-centered approach from the decision for surgery through recovery. By providing consistent information and predictable milestones, these programs reduce anxiety and uncertainty and are associated with high patient satisfaction when communication and education are delivered effectively [11].

Outpatient arthroplasty pathways extend this patient-centered ethos with rigorous selection criteria, structured education on pain, mobility, and exercises, and robust follow-up. Peacock et al [12] show that such pathways sustain high-quality care while aligning with cost-reduction goals; patients in the pilot phase reported overall satisfaction with the care process.

Finally, virtual clinics and digital adjuncts strengthen follow-up quality. A standardized combination of patient-reported questionnaires and radiology reports helps identify patients who need in-person assessment, streamlining follow-up, reducing unnecessary visits, and improving satisfaction [14]. In parallel, patient infotainment systems have been associated with improved care processes—such as fewer medical orders while maintaining medical quality—likely by enhancing education, setting expectations, and promoting patient autonomy in care management [15].

Resource Management

Clinical pathways in orthopedic surgery help reduce health care costs, primarily due to reduced complications and shorter hospital stays. Enhanced recovery protocols facilitate earlier discharges, which potentially decrease overall treatment costs and free inpatient capacity for other cases [6,28]. Additionally, a comparison study revealed that shortened hospital stays could lead to savings in subacute care, amounting to significant financial efficiencies [27]. Maintaining low readmission rates alongside shorter length of stay has further supported cost-efficiency, particularly in high-volume total joint arthroplasty (TJA) practices [22].

Within ICPs, economic findings have been mixed, several reports indicate substantial cost savings, while others highlight context-dependent variability and the need for further optimization [7]. Process improvements aligned with pathway implementation, such as Lean Six Sigma, have also been associated with more efficient bed turnover and reduced direct costs, reinforcing the operational value of standardized care processes [29].

Fast-track programs contribute to cost efficiency by promoting earlier discharge readiness and stable perioperative flow, helping services manage higher volumes without additional inpatient resource strain. In settings where low readmission rates are maintained, these programs reinforce the financial advantages of rapid recovery systems and minimize perioperative complications that drive costs [30].

The adoption of outpatient TJA pathways is increasingly recognized for its cost-effectiveness, producing substantial savings for both providers and patients through same-day discharge of carefully selected cases and reduced reliance on inpatient resources [12]. At the episode level, pathway implementation in total knee arthroplasty (TKA) has been associated with a mean cost decrease of about US $1252 per surgery, underscoring tangible financial benefits of standardized care bundles [8]. More broadly, the economic impact of defined care pathways includes reductions in hospital stay length and the use of post-acute services, with substantial cost differentials per episode [8].

Virtual clinics introduce additional savings by minimizing routine face-to-face follow-ups in hip and knee arthroplasty, optimizing the allocation of clinical resources and lowering the direct costs of outpatient services [13,14]. Beyond the index admission, preoperative physical therapy, often embedded within pathway frameworks, has been associated with decreased post-acute care use and related costs, further contributing to overall health care savings across the episode of care [27].

Sustainability

Sustainability in arthroplasty relates to preserving system capacity and quality over time, economically, operationally, environmentally (by proxy), and in terms of equitable access, rather than only achieving short-term episode efficiency. Across CPW subtypes, sustainability signals were most evident, where earlier discharge and standardized transitions translated into fewer bed-days per 100 cases, steadier workflows, and avoidable follow-ups without a safety penalty.

Within ERAS, durable capacity gains derive from consistent reductions in bed-days at scale, with programs reporting earlier discharge and no increase in short-term complications or readmissions. These effects support predictable staffing, theater and ward use and are accompanied in several programs by lower episode costs, indicating economic sustainability beyond single-episode efficiency [6,8]. Environmentally, ERAS-driven bed-days avoided serve as a pragmatic proxy for reduced energy and water use at ward level, while standardized escalation criteria help preserve equity by ensuring timely access to higher-intensity care when needed.

ICPs contribute to sustainability by coordinating care across settings, reducing unwarranted variability and unplanned care and making workforce demand more predictable. Multisite implementations describe shorter length of stay (LOS) with maintained safety and generally favorable, though context-dependent, cost signals, consistent with economic and operational sustainability. Standardized transitions and bundle adherence also support continuity and equity, particularly when patient education and criteria for escalation are explicit [24].

In fast-track models, large cohorts demonstrate marked LOS reductions without increased readmissions, enabling services to expand activity while maintaining safety. The sustained drop in bed-days per 100 cases, coupled with earlier mobilization and discharge readiness, stabilizes perioperative flow and staffing needs, key features of workforce and operational sustainability. These effects also mitigate downstream use of post-acute services where pathways incorporate robust discharge planning and community coordination [22].

Outpatient arthroplasty pathways reconfigure care by shifting low-risk cases to same-day discharge under clear selection and escalation criteria. This preserves inpatient capacity for higher-acuity patients and, in successful implementations, lowers episode costs, aligning economic and operational sustainability with maintained access and patient experience (ie, equity). Where shorter admissions replace longer stays, comparative analyses suggest savings in subacute care, further reinforcing system-level benefits [8,12,23,31].

Finally, virtual clinics deliver a proportion of follow-ups remotely using protocolized questionnaires and standardized radiology reports, preserving access and triage while reducing in-person visits. This approach smoothens outpatient workloads (operational sustainability) and introduces an environmental proxy benefit via travel avoided, provided escalation criteria are well defined and safety signals remain acceptable. Programs report optimized clinic capacity and reduced direct outpatient costs, consistent with sustained service delivery at scale [7].

To consolidate the findings of our review, we mapped each CPW subtype against the three core domains explored, sustainability, quality, and resource management in Table 2. This comparative summary allows for a clearer understanding of how each model contributes across these dimensions.

Table . CPW subtypes by Sustainability, Quality, and Resource Management.

CPW subtype	Quality	Resource management	Sustainability
ERAS (Enhanced Recovery After Surgery)	• High patient satisfaction and adherence to bundles; improved functional recovery in ERAS settings [24,28].	• Earlier discharge and fewer bed-days per case; lower episode cost reported in pathway-based programs [8,11].	• Discharge within 0 - 3 days without increased morbidity/morbidity/readmissions; updated ERAS protocols further reduce LOS [6,24] • Accelerated functional recovery and shorter hospital stays [11] • Reduced hospitalization times with updated ERAS protocols [6,28]
ICP (Integrated Clinical Pathway)	• Reduces variability and ensures high-quality care in elective surgeries [7,27] • Improves postoperative care and reduces complications [24] - Enhances adherence and satisfaction [7] • Preoperative physical therapy reduces post-acute care use by 29% [27]	• Subacute care savings from shorter stays [32]	• Reduced hospital stay for lower limb surgeries [24] • LOS reduction from 6.3 to 4.9 days, no increased readmissions [7] • Standardization reduces recovery time and resource use [13]
Fast-track Pathway	• High patient satisfaction with person-centered care [11,32]	• Supports cost-efficiency in high-volume TJA practices [8]	• Reduced stress and promoted early mobilization and recovery [22] • Nurse-led care enhances efficiency and reduces per-patient resource use [12]
Outpatient Arthroplasty Pathway	• High-quality care and strong patient satisfaction [12,23]	• Post-acute care cost savings per episode [13]	• 84.6% same-day discharge success rate [23]
Virtual Clinic	• Improves satisfaction via education and autonomy [22] • Streamlined follow-up using questionnaires and radiology [5]	• Mixed but promising cost reduction results [24]	• - Reduced hospital stay for TKA [14] • - Frees surgeon time by triaging unnecessary follow-ups [5]

This table compares five CPW subtypes—ERAS, ICP, Fast-track, Outpatient Arthroplasty, and Virtual Clinic—across three key dimensions: Sustainability, Quality, and Resource Management. ERAS promotes early discharge (0 - 3 d), improves recovery and patient satisfaction, and reduces hospital and postoperative costs. ICP standardizes care to lower hospital stays and complications, enhances adherence to best practices, and provides cost savings with Lean process improvements. Fast-track supports early mobilization and nurse-led care, delivering high patient satisfaction while reducing readmissions and resource use. Outpatient Arthroplasty achieves high same-day discharge rates, ensures quality in low-risk surgeries, and significantly lowers financial burdens. Virtual Clinics enhance sustainability via remote follow-up, improve patient autonomy and satisfaction, and reduce outpatient and complication-related costs.

Discussion

Principal Findings

This review found that CPWs in hip and knee arthroplasty consistently improve near-term efficiency and quality while generating system-level sustainability gains, fewer bed-days per 100 cases, steadier perioperative workflows, and avoidable follow-up reductions. Benefits were observed across ERAS, ICPs, fast-track, outpatient pathways, and virtual clinics, with no apparent trade-off in short-term safety in appropriately selected cohorts.

Indirect Outcomes

CPWs not only improve direct clinical outcomes but also lead to several indirect benefits. For example, the increased early mobilization and avoidance of continuous urinary catheters associated with CPWs have been linked to fewer complications and decreased readmission rates. Furthermore, this is underscored by the “halo effect” seen with care pathways, whereby improvements in one procedure translate to gains in others [26]. These indirect benefits highlight the comprehensive impact of CPWs on overall health care quality.

Applicability of Clinical Pathways

The applicability of CPWs in orthopedics extends beyond immediate postoperative care. These pathways are crucial for addressing common problems encountered by patients, such as

confusion about post-discharge medication and difficulty in assessing wound healing [31]. The integration of digital tools within CPWs can enhance patient education and engagement, providing clear guidance and support throughout the patient's journey [30]. This approach not only improves patient satisfaction but also reduces the burden on health care providers.

Importance of Initiating and Improving CPWs

Initiating and continuously improving CPWs is vital for maintaining and enhancing their effectiveness. The development of a standardized virtual clinic model for the follow-up of hip and knee arthroplasty patients, for instance, represents a significant innovation in orthopedic care delivery [14]. By leveraging expert consensus, this model aims to maintain the quality of care while reducing the resource burden of traditional follow-up methods, leading to enhanced patient satisfaction and cost efficiencies.

Digital Potential in Clinical Pathways

The integration of digital tools in CPWs has the potential to revolutionize orthopedic care. Digital tools can improve synchronization across different health care settings, enhancing disease management and data analysis capabilities. This digital evolution supports informed decision-making and efficient resource management, resulting in cost savings for health care systems [27]. The use of patient infotainment systems, as demonstrated by Huang et al [15], can reduce hospital stays without compromising care quality, highlighting the potential of digital solutions in enhancing CPWs.

Future clinical pathways should be digitally integrated and informed by policy and incentive frameworks—such as the Health Information Technology for Economic and Clinical Health Act and the Centers for Medicare & Medicaid Services' Promoting Interoperability program—with platform designs that support seamless integration with the electronic health record and efficient clinician review of patient-generated health data and patient-reported outcomes [16]. This adaptability will be critical to addressing rising care demands, workforce shortages, and maintaining high-quality, patient-centered outcomes [9].

Clinical pathways are key enablers of safe, efficient, and sustainable arthroplasty care. Standardized sustainability metrics and thoughtful integration of digital tools will be essential to maintain quality while expanding capacity for rising arthroplasty demand.

Challenges and Future Directions

Despite the benefits, implementing CPWs also presents challenges. The tension between adhering to fast-track requirements and addressing individual patient needs highlights the complexities of maintaining quality while standardizing care [33]. Future efforts should focus on refining CPWs to balance standardization with personalized care, leveraging digital advancements to support these goals.

The rapid implementation of outpatient TJA pathways during the COVID-19 pandemic is an example of how CPWs can adapt to emerging challenges. This initiative by Peacock et al [12] demonstrated the feasibility of outpatient surgeries to alleviate the strain on inpatient resources while maintaining high standards of patient care and satisfaction. Such adaptive strategies provide valuable insights for surgical centers facing similar challenges, emphasizing the importance of flexibility and innovation in CPWs.

The evolution of CPWs must align with new health care policies emphasizing value-based care and patient-reported outcomes (PROMs). In the United States, models like TEAM (Transforming Episode Accountability Model) and BPCI (Bundled Payments for Care Improvement) link reimbursement to PROMs (eg, HOOS Jr., KOOS Jr.), requiring CPWs to integrate these tools to remain compliant and improve care quality [2,34].

In Europe, initiatives like the Santeon Group in the Netherlands, Sweden's Quality Registries, and Germany's DiGA program demonstrate how digital and outcome-driven frameworks are being institutionalized [10,34,35]. The Horizon Europe agenda supports broader adoption through interoperability, PROM integration, and digital innovation [36].

A cross-national review by Srivastava et al [37] emphasizes key enablers for successful digital CPWs: coordinated governance, reimbursement models, workforce training, data sharing, and patient involvement. Countries such as Finland and the United Kingdom illustrate both progress and challenges, especially in IT integration and equitable access [38].

Conclusion

This review underscores the essential role of clinical pathways in orthopedic surgery, particularly in knee and hip arthroplasties. Through standardizing patient care, optimizing resource use, and enhancing sustainability, CPWs significantly improve patient outcomes, reduce variability in clinical practice, and contribute to greater health care efficiency. The integration of digital tools, such as virtual clinics and patient engagement platforms, further amplifies these benefits by facilitating remote follow-up, improving patient satisfaction, and reducing unnecessary hospital visits. However, challenges persist in balancing standardized protocols with individualized patient care. Future research should focus on refining CPWs and aligning them with global health care policies.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Checklist 1
PRISMA 2020 checklist.
[DOCX File, 276 KB - periop_v8i1e78174_app1.docx]

References

1. Kurtz S, Ong K, Lau E, Mowat F, Halpern M. Projections of primary and revision hip and knee arthroplasty in the United States from 2005 to 2030. J Bone Joint Surg Am 2007 Apr;89(4):780-785. [doi: 10.2106/JBJS.F.00222] [Medline: 17403800]
2. Navathe AS, Troxel AB, Liao JM, et al. Cost of joint replacement using bundled payment models. JAMA Intern Med 2017 Feb 1;177(2):214-222. [doi: 10.1001/jamainternmed.2016.8263] [Medline: 28055062]
3. Ayalon O, Liu S, Flics S, Cahill J, Juliano K, Cornell CN. A multimodal clinical pathway can reduce length of stay after total knee arthroplasty. HSS J 2011 Feb;7(1):9-15. [doi: 10.1007/s11420-010-9164-1] [Medline: 22294952]
4. Rotter T, Kinsman LD, Alsius A, et al. Clinical pathways for secondary care and the effects on professional practice, patient outcomes, length of stay and hospital costs. Cochrane Database Syst Rev 2025 May 14;5(5):CD006632. [doi: 10.1002/14651858.CD006632.pub3] [Medline: 40365866]
5. Foni NO, Costa LAV, Paião ID, et al. Clinical pathway improves medical practice in total knee arthroplasty. PLoS ONE 2020;15(5):e0232881. [doi: 10.1371/journal.pone.0232881] [Medline: 32379840]
6. Auyong DB, Allen CJ, Pahang JA, Clabeaux JJ, MacDonald KM, Hanson NA. Reduced length of hospitalization in primary total knee arthroplasty patients using an updated Enhanced Recovery After Orthopedic Surgery (ERAS) pathway. J Arthroplasty 2015 Oct;30(10):1705-1709. [doi: 10.1016/j.arth.2015.05.007] [Medline: 26024988]
7. Zhou H, Ngune I, Roberts PA, Della PR. Integrated clinical pathways for lower limb orthopaedic surgeries: an updated systematic review. J Clin Nurs 2023 Jun;32(11-12):2433-2454. [doi: 10.1111/jocn.16344] [Medline: 35703679]
8. Tessier JE, Rupp G, Gera JT, DeHart ML, Kowalik TD, Duwelius PJ. Physicians with defined clear care pathways have better discharge disposition and lower cost. J Arthroplasty 2016 Sep;31(9 Suppl):54-58. [doi: 10.1016/j.arth.2016.05.001] [Medline: 27329578]
9. Porter ME, Teisberg EO. Redefining Health Care: Creating Value-Based Competition on Results: Harvard Business School Press; 2006.
10. van Hoorn ES, Ye L, van Leeuwen N, Raat H, Lingsma HF. Value-based integrated care: a systematic literature review. Int J Health Policy Manag 2024;13:8038. [doi: 10.34172/ijhpm.2024.8038] [Medline: 38618830]
11. Berg U, Berg M, Rolfson O, Erichsen-Andersson A. Fast-track program of elective joint replacement in hip and knee-patients' experiences of the clinical pathway and care process. J Orthop Surg Res 2019 Jun 21;14(1):186. [doi: 10.1186/s13018-019-1232-8] [Medline: 31227003]
12. Peacock S, Wolfstadt J, Peer M, Gleicher Y. Rapid implementation of an outpatient arthroplasty care pathway: a COVID-19-driven quality improvement initiative. BMJ Open Qual 2022 Mar;11(1):e001698. [doi: 10.1136/bmjoq-2021-001698] [Medline: 35318244]
13. Greenhalgh T, Vijayaraghavan S, Wherton J, et al. Virtual online consultations: advantages and limitations (VOCAL) study. BMJ Open 2016 Jan 29;6(1):e009388. [doi: 10.1136/bmjopen-2015-009388] [Medline: 26826147]
14. Preston N, McHugh GA, Hensor EMA, et al. Developing a standardized approach to virtual clinic follow-up of hip and knee arthroplasty. Bone Joint J 2019 Aug;101-B(8):951-959. [doi: 10.1302/0301-620X.101B8.BJJ-2018-1566.R1] [Medline: 31362551]
15. Huang S, Kuo ML, Yu HM, et al. Clinical information and guidance shared via a patient infotainment system can reduce hospital stay and maintain 2 medical quality for total knee arthroplasty: a single-blinded quasi-randomised controlled trial. Int J Nurs Stud 2020 Apr;104:103440. [doi: 10.1016/j.ijnurstu.2019.103440] [Medline: 32105971]
16. Lordon RJ, Mikles SP, Kneale L, et al. How patient-generated health data and patient-reported outcomes affect patient–clinician relationships: a systematic review. Health Informatics J 2020 Dec;26(4):2689-2706. [doi: 10.1177/1460458220928184] [Medline: 32567460]
17. Beaupre LA, Jones CA, Saunders LD, Johnston DWC, Buckingham J, Majumdar SR. Best practices for elderly hip fracture patients. A systematic overview of the evidence. J Gen Intern Med 2005 Nov;20(11):1019-1025. [doi: 10.1111/j.1525-1497.2005.00219.x] [Medline: 16307627]
18. Campagner A, Milella F, Guida S, Bernareggi S, Banfi G, Cabitza F. Assessment of fast-track pathway in hip and knee replacement surgery by propensity score matching on patient-reported outcomes. Diagnostics (Basel) 2023 Mar 21;13(6):1189. [doi: 10.3390/diagnostics13061189] [Medline: 36980497]
19. Jain SN, Lamture Y, Krishna M. Enhanced recovery after surgery: exploring the advances and strategies. Cureus 2023 Oct;15(10):e47237. [doi: 10.7759/cureus.47237] [Medline: 38022245]
20. Kehlet H. Fast-track surgery-an update on physiological care principles to enhance recovery. Langenbecks Arch Surg 2011 Jun;396(5):585-590. [doi: 10.1007/s00423-011-0790-y] [Medline: 21468643]
21. Campbell H, Hotchkiss R, Bradshaw N, Porteous M. Integrated care pathways. BMJ 1998 Jan 10;316(7125):133-137. [doi: 10.1136/bmj.316.7125.133] [Medline: 9462322]

22. Petersen PB, Kehlet H, Jørgensen CC, Lundbeck Foundation Centre for Fast-track Hip and Knee Replacement Collaborative Group. Improvement in fast-track hip and knee arthroplasty: a prospective multicentre study of 36,935 procedures from 2010 to 2017. Sci Rep 2020 Dec 4;10(1):21233. [doi: 10.1038/s41598-020-77127-6] [Medline: 33277508]
23. Berger RA, Sanders SA, Thill ES, Sporer SM, Della Valle C. Newer anesthesia and rehabilitation protocols enable outpatient hip replacement in selected patients. Clin Orthop Relat Res 2009 Jun;467(6):1424-1430. [doi: 10.1007/s11999-009-0741-x] [Medline: 19252961]
24. Wainwright TW, Gill M, McDonald DA, et al. Consensus statement for perioperative care in total hip replacement and total knee replacement surgery: Enhanced Recovery After Surgery (ERAS®) Society recommendations. Acta Orthop 2020 Jun;91(3):3-19. [doi: 10.1080/17453674.2020.1724674] [Medline: 32056486]
25. Moher D, Shamseer L, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015 Jan 1;4(1):1. [doi: 10.1186/2046-4053-4-1] [Medline: 25554246]
26. Barber C, Fraser JF, Mendez GG, Bradley B, Loftus TJ, Jacofsky DJ. The halo effect: an unintended benefit of care pathways. J Knee Surg 2017 Mar;30(3):264-268. [doi: 10.1055/s-0036-1584577] [Medline: 27362923]
27. Snow R, Granata J, Ruhil AVS, Vogel K, McShane M, Wasielewski R. Associations between preoperative physical therapy and post-acute care utilization patterns and cost in total joint replacement. J Bone Joint Surg Am 2014 Oct 1;96(19):e165. [doi: 10.2106/JBJS.M.01285] [Medline: 25274793]
28. Didden AGM, Punt IM, Feczko PZ, Lenssen AF. Enhanced recovery in usual health care improves functional recovery after total knee arthroplasty. Int J Orthop Trauma Nurs 2019 Aug;34:9-15. [doi: 10.1016/j.ijotn.2019.03.003] [Medline: 31272919]
29. Improta G, Balato G, Romano M, et al. Lean Six Sigma: a new approach to the management of patients undergoing prosthetic hip replacement surgery. J Eval Clin Pract 2015 Aug;21(4):662-672. [doi: 10.1111/jep.12361] [Medline: 25958776]
30. Yamashita T, Wakata Y, Hamai S, et al. Temporal relation extraction in outcome variances of clinical pathways. Stud Health Technol Inform 2015;216:1077. [doi: 10.3233/978-1-61499-667-5-1077] [Medline: 26262376]
31. Hägglund M, Bolin P, Koch S. Experiences as input to eHealth design - a hip surgery patient journey case. Stud Health Technol Inform 2015;210:672-674. [doi: 10.3233/978-1-61499-512-8-672] [Medline: 25991235]
32. Berthelsen CB, Frederiksen K. Orchestrating care through the fast-track perspective: a qualitative content analysis of the provision of individualised nursing care in orthopaedic fast-track programmes. Int J Orthop Trauma Nurs 2017 Feb;24:40-49. [doi: 10.1016/j.ijotn.2016.04.006] [Medline: 27615119]
33. Schotanus MGM, Bemelmans YFL, Grimm B, Heyligers IC, Kort NP. Physical activity after outpatient surgery and enhanced recovery for total knee arthroplasty. Knee Surg Sports Traumatol Arthrosc 2017 Nov;25(11):3366-3371. [doi: 10.1007/s00167-016-4256-1] [Medline: 27492381]
34. Hoque DME, Kumari V, Hoque M, Ruseckaite R, Romero L, Evans SM. Impact of clinical registries on quality of patient care and clinical outcomes: a systematic review. PLoS ONE 2017;12(9):e0183667. [doi: 10.1371/journal.pone.0183667] [Medline: 28886607]
35. Gerke S, Stern AD, Minssen T. Germany's digital health reforms in the COVID-19 era: lessons and opportunities for other countries. NPJ Digit Med 2020;3:94. [doi: 10.1038/s41746-020-0306-7] [Medline: 32685700]
36. Horizon Europe Strategic Plan 2021–2024: Health Cluster. Brussels: European Commission. 2021. URL: https://research-and-innovation.ec.europa.eu [accessed 2025-06-28]
37. Srivastava D, Van Kessel R, Delgrange M, Cherla A, Sood H, Mossialos E. A framework for digital health policy: insights from virtual primary care systems across five nations. PLOS Digit Health 2023 Nov;2(11):e0000382. [doi: 10.1371/journal.pdig.0000382] [Medline: 37939131]
38. Briggs TWR. A national review of adult elective orthopaedic services in England: getting it right first time. : NHS England; 2015 URL: https://gettingitrightfirsttime.co.uk/wp-content/uploads/2018/07/GIRFT-National-Report-Mar15-Web.pdf [accessed 2025-06-28]

Abbreviations

CPW: Clinical Pathway
ERAS: Enhanced Recovery After Surgery
ICPs: Integrated Clinical Pathways
LOS: length of stay
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROM: Patient-Reported Outcome Measure
TJA: total joint arthroplasty
TKA: total knee arthroplasty

Edited by J Shiffermiller; submitted 28.05.25; peer-reviewed by C Deans, U Modebelu; revised version received 01.09.25; accepted 08.09.25; published 14.10.25.

Please cite as:
Godinho M, Mações F, Gonçalves H, Silva F
Clinical Pathways in Knee and Hip Arthroplasty: Narrative Review on Sustainability, Quality, and Resource Management
JMIR Perioper Med 2025;8:e78174
URL: https://periop.jmir.org/2025/1/e78174
doi:10.2196/78174

Corrigenda and Addenda

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

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

[1]Department of Public and Occupational Health, Amsterdam UMC location Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
[2]Societal Participation & Health, Amsterdam Public Health research institute, Amsterdam, The Netherlands
[3]Department of Public and Occupational Health, Amsterdam UMC location University of Amsterdam, Amsterdam, The Netherlands
[4]Department of Obstetrics and Gynaecology, Amsterdam UMC location 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

Related Article:

Correction of: https://periop.jmir.org/2025/1/e58878

(JMIR Perioper Med 2025;8:e71874) doi:10.2196/71874

In "A Patient-Oriented Implementation Strategy for a Perioperative mHealth Intervention: Feasibility Cohort Study" JMIR Perioper Med 2025;8:e58878) the authors noted one error.

The affiliations were incorrectly published as:

> *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*

These have been revised as follows:

> *Daan Toben[1,2], MSc; Astrid de Wind[2,3], PhD; Eva van der Meij[2,3], PhD; Judith A F Huirne[2,4], PhD; Johannes R Anema[1,2], PhD*
>
> *[1]Department of Public and Occupational Health, Amsterdam UMC location Vrije Universiteit Amsterdam, Amsterdam, The Netherlands*
>
> *[2]Societal Participation & Health, Amsterdam Public Health research institute, Amsterdam, The Netherlands*
>
> *[3]Department of Public and Occupational Health, Amsterdam UMC location University of Amsterdam, Amsterdam, The Netherlands*
>
> *[4]Department of Obstetrics and Gynaecology, Amsterdam UMC location University of Amsterdam, Amsterdam, The Netherlands*

The correction will appear in the online version of the paper on the JMIR Publications website on February 12, 2025 together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 28.01.25; this is a non–peer-reviewed article;accepted 28.01.25; published 12.02.25.

Please cite as:
Toben D, de Wind A, van der Meij E, Huirne JAF, Anema JR
Correction: A Patient-Oriented Implementation Strategy for a Perioperative mHealth Intervention: Feasibility Cohort Study
JMIR Perioper Med 2025;8:e71874
URL: https://periop.jmir.org/2025/1/e71874
doi:10.2196/71874
PMID:

Corrigenda and Addenda

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

Related Article:

Correction of: https://periop.jmir.org/2023/1/e34453

(JMIR Perioper Med 2025;8:e84392) doi:10.2196/84392

Following the publication of "Patient Safety of Perioperative Medication Through the Lens of Digital Health and Artificial Intelligence" [1], concerns were raised regarding the corresponding author Jiancheng Ye's rate of self-citation within the article's reference list. Further investigation into this matter identified that numerous self-citations were added to the manuscript following acceptance of the article without adequate disclosure.

The author was contacted for a response regarding this matter. The author indicated that the added references were pertinent to the article and its findings, and provided documentation justifying the inclusion of the identified self-citations.

Following review of this response, the JMIR Publications Editorial Office determined that several of the added references were not relevant to the article or were used to support a generic statement where another reference could have been used instead of self-citation. Therefore, the JMIR Publications Editorial Office is issuing this corrigendum to remove the following references:

- Reference 5: Feinglass J, Wang JA, Ye J, Tessier R, Kim H. Hospital care for opioid use in Illinois, 2016-2019. J Behav Health Serv Res 2021 Oct;48(4):597-609 [doi: 10.1007/s11414-020-09748-8] [Medline: 33502670]
- Reference 6: Ye J. Health information system's responses to COVID-19 pandemic in China: a national cross-sectional study. Appl Clin Inform 2021 Mar 19;12(2):399-406 [doi: 10.1055/s-0041-1728770] [Medline: 34010976]
- Reference 19: Ye J, Orji IA, Baldridge AS, Ojo TM, Shedul G, Ugwuneji EN, Hypertension Treatment in Nigeria Program Investigators. Characteristics and patterns of retention in hypertension care in primary care settings from the Hypertension Treatment in Nigeria program. JAMA Netw Open 2022 Sep 01;5(9):e2230025 [doi: 10.1001/jamanetworkopen.2022.30025] [Medline: 36066896]
- Reference 21: Ye J, Ren Z. Examining the impact of sex differences and the COVID-19 pandemic on health and health care: findings from a national cross-sectional study. JAMIA Open 2022 Oct;5(3):ooac076 [doi: 10.1093/jamiaopen/ooac076] [Medline: 36177395]
- Reference 24: Ye J, Ma Q. The effects and patterns among mobile health, social determinants, and physical activity: a nationally representative cross-sectional study. AMIA Jt Summits Transl Sci Proc 2021;2021:653-662 [Medline: 34457181]
- Reference 30: Li Q, Fu R, Zhang J, Wang R, Ye J, Xue N, et al. Label-free method using a weighted-phase algorithm to quantitate nanoscale interactions between molecules on DNA microarrays. Anal Chem 2017 Mar 21;89(6):3501-3507 [doi: 10.1021/acs.analchem.6b04596] [Medline: 28230978]

The JMIR Publications Editorial Office regrets that these issues were not identified prior to publication.

The correction will appear in the online version of the paper on the JMIR Publications website together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Reference

1. Ye J. Patient safety of perioperative medication through the lens of digital health and artificial intelligence. JMIR Perioper Med 2023 May 31;6:e34453 [FREE Full text] [doi: 10.2196/34453] [Medline: 37256663]

Submitted 18.09.25; this is a non–peer-reviewed article;accepted 18.09.25; published 08.10.25.

Please cite as:
Correction: Patient Safety of Perioperative Medication Through the Lens of Digital Health and Artificial Intelligence
JMIR Perioper Med 2025;8:e84392
URL: https://periop.jmir.org/2025/1/e84392
doi:10.2196/84392
PMID:

Preoperative Anxiety and Information Desire Among Patients Undergoing Elective Surgery in Northern Sudan: Multicenter Cross-Sectional Study

Abeer Ahmed[1*], MBBS; Mohamed Nasur[2]; Eman Mohamed[3]; Amna Faragalla[4]; Mustafa Ahmed[5]; Murouj Mohammed[6], MBBS; Abdinur Yusuf[2], MBBS; Mohamed Issak[2*], MBBS

[1]Department of Public Health, Postgraduate School, University of Dongola, Dongola, Sudan

[2]Faculty of Medicine and Health Science, University of Dongola, Faculty of Medicine Street, Dongola, Sudan

[3]Faculty of Medicine, University of Gezira, Wad Madani, Sudan

[4]Department of Internal Medicine, Al-Wifaq Al-Qatari Hospital, Al Dabbah, Sudan

[5]Faculty of Medicine, Merowe University of Technology, Merowe, Sudan

[6]Internal Medicine Council, Sudan Medical Specialization Board, Atbara, Sudan

[*]these authors contributed equally

Corresponding Author:
Mohamed Issak, MBBS
Faculty of Medicine and Health Science, University of Dongola, Faculty of Medicine Street, Dongola, Sudan

Abstract

Background: Preoperative anxiety is a common psychological condition, and many patients express a desire for more information before surgery. Understanding the prevalence and associated factors of both preoperative anxiety and the desire for information can improve patient care.

Objective: This study aimed to assess the prevalence of preoperative anxiety and desire for information, as well as their associated sociodemographic, medical, and surgical factors, among patients undergoing elective surgery in Northern State, Sudan.

Methods: A hospital-based, multicenter, cross-sectional study was conducted from November 2024 to February 2025 in Northern State, Sudan, involving patients undergoing elective surgery. Data were collected through face-to-face interviews using a structured questionnaire and the validated Arabic version of the Amsterdam Preoperative Anxiety and Information Scale (APAIS). Chi-square tests and univariate and multivariate logistic regression were performed to identify the associated factors and the magnitude, with statistical significance set at $P<.05$.

Results: Of the 380 patients approached, 305 participated in the study (response rate=80.3%): 173 of the 305 participants (56.7%) were male, and the median age was 43 (IQR 30 - 64) years. Most participants were married (n=207, 67.9%), educated (n=248, 81.3%), and had family support (n=253, 83.0%). Regarding surgical characteristics, the majority underwent either intermediate (n=136, 44.6%) or major (n=142, 46.6%) procedures. General anesthesia was the most common type used (n=159, 52.2%), and most participants (n=169, 55.4%) underwent surgery in public hospitals. Most participants reported that their surgeries were not covered by insurance (n=264, 86.6%) and described good sleep quality the night before surgery (n=221, 72.5%). Of the 305 participants, 75 (24.6%) experienced preoperative anxiety, whereas 92 (30.1%) expressed a moderate to high desire for information. Preoperative anxiety was significantly associated with family support (adjusted odds ratio [aOR] 7.12, 95% CI 2.64 - 19.23; $P<.001$), surgery in public hospitals (aOR 4.31, 95% CI 2.30 - 8.07; $P<.001$), poor sleep quality the night before surgery (aOR 2.85, 95% CI 1.51 - 5.38; $P=.001$), and American Society of Anesthesiologists (ASA) classification III/IV (aOR 2.36, 95% CI 1.00 - 5.54; $P=.049$). Similarly, a higher desire for information was significantly associated with being educated (aOR 2.48, 95% CI 1.00 - 6.11; $P=.049$), having family support (aOR 4.10, 95% CI 1.81 - 9.30; $P=.001$), undergoing surgery in a public hospital (aOR 3.57, 95% CI 1.93 - 6.61; $P<.001$), and being classified as ASA III/IV (aOR 3.26, 95% CI 1.39 - 7.64; $P=.001$).

Conclusions: Preoperative anxiety and desire for information are common among Sudanese patients. Family involvement may paradoxically increase anxiety and the desire for more information due to shared concerns and cultural factors. Other significant predictors of anxiety include poor sleep quality and higher ASA classification. Additionally, education, family support, and chronic diseases were associated with a higher desire for information. Addressing these factors may alleviate preoperative anxiety, satisfy communication needs, and improve preoperative care.

(JMIR Perioper Med 2025;8:e75736) doi:10.2196/75736

KEYWORDS

preoperative anxiety; desire for information; risk factors; elective surgery; Sudan

Introduction

Anxiety is a subjective state of emotional unease, distress, or apprehension, often arising from the anticipation of danger and accompanied by autonomic and somatic symptoms that can impair functioning. While occasional anxiety is common, anxiety disorders involve intense and excessive fear and worry [1]. Approximately 33.7% of the global population experiences an anxiety disorder during their lifetime [2].

Anxiety becomes particularly concerning in medical contexts, such as surgery, where it can significantly impact outcomes and well-being [3]. Elective surgical procedures, which are nonemergency procedures that can be delayed for at least 24 hours, often trigger preoperative anxiety in patients [4]. This anxiety is characterized by worry, nervousness, and unease about the impending procedure and its potential outcomes [3,5].

Preoperative anxiety may have significant negative impacts on patients. It is associated with increased anesthesia requirements and hemodynamic instability during surgery as well as postoperative complications such as nausea, vomiting, and dizziness [6,7]. High anxiety levels are also linked to poor postoperative pain control, increased analgesic needs, delayed recovery, longer hospital stays, and complications such as surgical site infections and impaired wound healing [8-10]. These adverse effects may reduce patient satisfaction, hinder rehabilitation, and increase health care costs, making preoperative anxiety not only a psychological issue but also a clinical and economic one.

The global pooled prevalence of preoperative anxiety is estimated at 48% [11]. In low- and middle-income countries (LMICs), the issue is exacerbated by limited health care resources, inadequate patient information, and cultural beliefs surrounding surgery. Studies in LMICs reveal even higher prevalence rates, ranging from 55% to 99%, highlighting the need for context-specific research and interventions [12]. Neighboring countries such as Ethiopia and Yemen also report varying rates, influenced by factors such as surgery type, patient demographics, prior surgical experiences, and individual coping mechanisms [13,14].

In Sudan, where the health care system faces significant constraints due to the ongoing war, the issue of preoperative anxiety among patients undergoing elective surgery may be particularly pronounced. Factors such as limited health facilities, lack of access to information, long waiting times, and communication barriers may contribute to heightened anxiety levels. Moreover, the socioeconomic conditions and cultural beliefs of the Sudanese population may shape patients' perceptions, expectations, and surgical experiences.

Despite growing global recognition of preoperative anxiety as a significant health concern, it remains underresearched in Sudan. Understanding this issue is essential for guiding health authorities and care providers in developing targeted interventions that reduce anxiety, enhance patient satisfaction, and improve surgical outcomes in resource-limited and conflict-affected settings such as Sudan.

Therefore, this pioneering study—the first of its kind in Sudan—aims to determine the prevalence of preoperative anxiety and desire for information and identify associated sociodemographic and clinical predictors among surgical patients in Northern Sudan.

Methods

Study Design

A multicenter, cross-sectional study was conducted over 4 months, from November 2024 to February 2025, in 4 hospitals with operation theaters in Northern State, Sudan: 2 public hospitals (Dongola Specialized Hospital and Karima Teaching Hospital) and 2 private hospitals (Al-Wifaq Al-Qatari Hospital and Al-Daman Hospital). The study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines for observational studies [15].

Study Population

The study included adult patients (≥18 years) undergoing elective surgery who provided written informed consent. Exclusion criteria were emergency or obstetric surgeries, a history of mental illness or cognitive impairment, and incomplete responses to the study questionnaire.

Sample Size and Sampling Technique

The sample size was calculated using Cochran's formula: sample size (n) = z2pqe2 [16]. Using a prevalence of 55.7% (from similar LMIC studies), 95% confidence level (z=1.96), and 5% margin of error, the minimum sample size was determined to be 380 [7,16]. In total, 305 of the 380 distributed surveys were completed, yielding a response rate of 80.3%. Patients were recruited using convenience sampling from surgical wards and outpatient clinics 1-2 days before their scheduled operation.

Data Collection

Data were collected through structured face-to-face interviews (within 48 h prior to the scheduled surgery) using a standardized questionnaire comprising 3 sections: a sociodemographic section with 7 items, a medical and surgical history section with 7 items, and the Amsterdam Preoperative Anxiety and Information Scale (APAIS) with 6 items .

The questionnaire was translated into Arabic, and the validated Arabic version of the APAIS was used [17]. A pilot test was conducted among 40 participants to ensure clarity and reliability, yielding a Cronbach α of 0.91, indicating excellent reliability [18]. The English version of the questionnaire is provided as Multimedia Appendix 1.

Measurement of Preoperative Anxiety and Desire for Information

The APAIS was used to assess preoperative anxiety and desire for information. It is a validated scale with 6 items, each scored

on a 5-point Likert scale ranging from 1 (“not at all”) to 5 (“extremely”). Preoperative anxiety was assessed through items 1, 2, 4, and 5, leading to a total score range of 4 - 20, with a score ≥11 indicating clinically significant anxiety. Desire for information was assessed using items 3 and 6. The total scores ranged from 2 to 10 and were categorized as low (2-4), average (5-7), or high (8-10) desire for information [19-21].

Data Analysis

Descriptive Analysis

SPSS (version 27; IBM Corp) was used for all statistical analyses. Descriptive statistics (frequencies and percentages) summarized the sociodemographic, medical, and surgical characteristics of participants, as well as the prevalence of preoperative anxiety and desire for information.

Bivariate Analysis

Chi-square tests were used to examine associations between independent variables (eg, sociodemographics and surgical details) and the outcomes: preoperative anxiety and desire for information. A *P* value of <.05 was considered statistically significant.

Multivariate Analysis

Variables with a *P* value <.05 in the bivariate analysis were included in a binary logistic regression model for preoperative anxiety and desire for information. Adjusted odds ratios (aORs) with 95% CI were calculated to determine independent predictors.

Ethical Considerations

The study was conducted in accordance with the Declaration of Helsinki. Ethical approval was obtained from the Research Ethical Committee of the Northern State Ministry of Health (REC-MoH-NS-41 - 2023) and the research offices of the involved hospitals. Written informed consent was obtained from all participants prior to their inclusion in the study. Participants were assured of the confidentiality of their responses, the anonymity of their data, and their right to withdraw from the study at any time without any impact on their medical care. No compensation was provided to participants for their involvement in the study.

Results

General Characteristics of the Study Population

A total of 380 patients undergoing elective surgery were approached, of whom 305 participated in this study (response rate=80.3%). Figure 1 shows the distribution of participants across the 4 participating hospitals. Approximately half of the participants were male (n=173, 56.7%) and aged between 31 and 65 years (n=161, 52.7%), with a median age of 43 (IQR 30 - 64) years. Most participants (n=207, 67.9%) were married, and the majority (n=248, 81.3%) had some level of education. Of the 305 participants, 114 (37.4%) were freelancers, and 117 (38.4%) were unemployed. Nearly half of the participants (n=151, 49.5%) had an average monthly income above US $250. Family support was reported by 83.0% of participants (253/305). The sociodemographic characteristics of the participants are summarized in Table 1.

Figure 1. Flowchart illustrating the distribution of study participants across 4 participating hospitals in Northern State, Sudan.

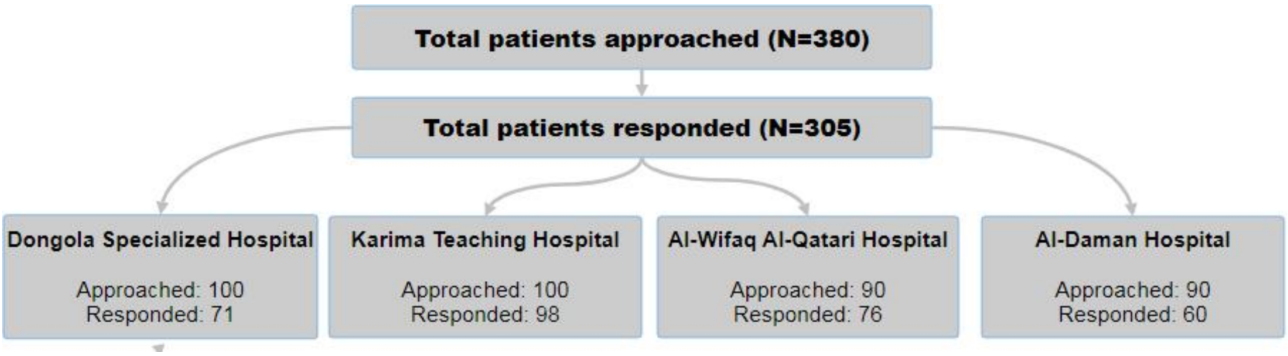


Table . Sociodemographic and medical/surgical characteristics of patients undergoing elective surgery at 4 participating hospitals in Northern State, Sudan, in 2024-2025 (N=305).

Characteristic	Participants, n (%)
Sex	
Male	173 (56.7)
Female	132 (43.3)
Age[a] (years)	
18 - 30	86 (28.2)
31 - 50	102 (33.4)
51 - 65	59 (19.3)
>65	58 (19.0)
Occupation	
Employed	46 (15.1)
Freelancer	114 (37.4)
Student	28 (9.2)
Unemployed	117 (38.4)
Marital status	
Married	207 (67.9)
Unmarried	98 (32.1)
Educational level	
Some level of education	248 (81.3)
No education	57 (18.7)
Average monthly income (US $)	
<50	125 (41.0)
50 - 250	29 (9.5)
>250	151 (49.5)
Family support	
Yes	253 (83.0)
No	52 (17.0)
Type of hospital	
Public hospital	169 (55.4)
Private hospital	136 (44.6)
ASA[b] classification	
ASA I and II	268 (87.9)
ASA III and IV	37 (12.1)
Type of surgery	
Minor	27 (8.9)
Intermediate	136 (44.6)
Major	142 (46.6)
Type of anesthesia	
General	160 (52.5)
Spinal	124 (40.7)
Local	21 (6.9)
Operation covered by insurance	

Characteristic	Participants, n (%)
Yes	41 (13.4)
No	264 (86.6)
Subjective sleep quality the night before the operation	
Good	221 (72.5)
Poor	84 (27.5)
Previous operations	
Yes	119 (39.0)
No	186 (61.0)
Chronic diseases	
Yes	106 (34.8)
No	199 (65.2)

[a]Median age 43 (IQR 30-64) years.

[b]ASA: American Society of Anesthesiologists.

Medical and Surgical Characteristics of the Study Population

The medical and surgical characteristics of the participants are detailed in Table 1. According to the American Society of Anesthesiologists (ASA) classification system, most patients (268/305, 87.9%) were classified as ASA I and II. Most patients underwent major (n=142, 46.6%) or intermediate (n=136, 44.6%) surgeries. General anesthesia was the most common type of anesthesia used (n=160, 52.5%), and most surgeries (n=169, 55.4%) were performed in public hospitals. Most surgeries (n=264, 86.6%) were not covered by insurance, and most patients (n=221, 72.5%) reported good subjective sleep quality the night before the operation. A significant proportion of patients had had no previous operations (n=186, 61.0%) and no chronic diseases (199, 65.2%).

Prevalence of Preoperative Anxiety and Desire for Information

The prevalence of preoperative anxiety and desire for information among the participants is shown in Tables 2 and 3. Overall, 75 of the 305 participants (24.6%) reported experiencing clinically significant preoperative anxiety. Of the 305 participants, 73 (23.9%) expressed an average desire for information, and 19 (6.2%) reported a high desire for information.

Table . Amsterdam Preoperative Anxiety and Information Scale (APAIS) scores of patients undergoing elective surgery at 4 participating hospitals in Northern State, Sudan, in 2024-2025 (N=305).

Statement	Not at all, n (%)	Slightly, n (%)	Moderately, n (%)	Quite a bit, n (%)	Extremely, n (%)
I am worried about the anesthesia	148 (48.5)	82 (26.9)	38 (12.5)	30 (9.8)	7 (2.3)
The anesthetic is on my mind continually	157 (54.5)	68 (22.3)	51 (16.7)	23 (7.5)	6 (2.0)
I would like to know as much as possible about the anesthesia	169 (55.4)	66 (21.6)	43 (14.1)	21 (6.9)	6 (2.0)
I am worried about the procedure	127 (41.6)	73 (23.9)	54 (17.7)	43 (14.1)	8 (2.6)
The procedure is on my mind continually	123 (40.3)	79 (25.9)	51 (16.7)	44 (14.4)	8 (2.6)
I would like to know as much as possible about the procedure	132 (43.3)	85 (27.9)	52 (17.0)	26 (8.5)	10 (3.3)

Table . Frequency of clinically significant preoperative anxiety and desire for information among patients undergoing elective surgery at 4 participating hospitals in Northern State, Sudan, in 2024-2025 (N=305).

Characteristic	Participants, n (%)
Clinically significant preoperative anxiety	
Present	75 (24.6)
Absent	230 (75.4)
Desire for information	
No desire	213 (69.8)
Average desire	73 (23.9)
High desire	19 (6.2)

Factors Associated With Preoperative Anxiety and Desire for Information

The factors associated with preoperative anxiety (Table 4) included the presence of family support (*P*=.02), poor subjective sleep quality the night before the operation (*P*=.01), surgery in a public hospital (*P*<.001), ASA classification III and IV (*P*=.02), and a history of chronic diseases (*P*=.05).

Table . Association of sociodemographic and medical/surgical characteristics with clinically significant preoperative anxiety among patients undergoing elective surgery at 4 participating hospitals in Northern State, Sudan, in 2024-2025 (N=305).

Characteristic	Clinically significant preoperative anxiety			
	Present, n (%)	Absent, n (%)	Chi-square (*df*)	*P* value
Sex			0.90 (1)	.34
Male	39 (22.5)	134 (77.5)		
Female	36 (27.3)	96 (72.7)		
Age (years)			2.05 (3)	.56
18 - 30	20 (23.3)	66 (76.7)		
31 - 50	30 (29.4)	72 (70.6)		
51 - 65	13 (22.0)	46 (78.0)		
>65	12 (20.7)	46 (79.3)		
Occupation			4.77 (3)	.19
Employed	13 (28.3)	33 (71.7)		
Freelancer	21 (18.4)	93 (81.6)		
Student	10 (35.7)	18 (64.3)		
Unemployed	31 (26.5)	86 (73.5)		
Marital status			0.29 (1)	.59
Married	49 (23.7)	158 (76.3)		
Single	26 (26.5)	72 (73.5)		
Educational level			1.88 (1)	.17
Some level of education	65 (26.2)	183 (73.8)		
No education	10 (17.5)	47 (82.5)		
Average monthly income (US $)			2.04 (2)	.36
<50	36 (28.8)	89 (71.2)		
50.00 - 250	33 (21.9)	118 (78.1)		
>250	6 (20.7)	23 (79.3)		
Family support			5.76 (1)	.02
Yes	69 (27.3)	184 (72.7)		
No	6 (11.5)	46 (88.5)		
Type of hospital			17.01 (1)	<.001
Public hospital	57 (33.7)	112 (66.3)		
Private hospital	18 (13.2)	118 (86.8)		
ASA[a] classification			5.78 (1)	.02
ASA I and II	60 (22.4)	208 (77.6)		
ASA III and IV	15 (40.5)	22 (59.5)		
Type of surgery			2.65 (2)	.27
Minor	9 (33.3)	18 (66.7)		
Intermediate	28 (20.6)	108 (79.4)		
Major	38 (26.8)	104 (73.2)		
Type of anesthesia			0.46 (2)	.79

Characteristic	Clinically significant preoperative anxiety			
	Present, n (%)	Absent, n (%)	Chi-square (*df*)	*P* value
General	37 (23.1)	123 (79.9)		
Spinal	32 (25.8)	92 (74.2)		
Local	6 (28.6)	15 (71.4)		
Operation covered by insurance			3.75 (1)	.11
Yes	6 (14.6)	35 (85.4)		
No	69 (26.1)	195 (73.9)		
Subjective sleep quality the night before the operation			7.74 (1)	.01
Good	45 (20.4)	176 (79.6)		
Poor	30 (35.7)	54 (64.3)		
Previous operations			1.35 (1)	.25
Yes	25 (21.0)	94 (79.0)		
No	50 (26.9)	136 (73.1)		
Chronic diseases			3.75 (1)	.05
Yes	33 (31.1)	73 (68.9)		
No	42 (21.1)	157 (78.9)		

[a]ASA: American Society of Anesthesiologists.

The factors associated with the desire for information (Table 5) included being employed (*P*=.01), being educated (*P*=.003), having family support (*P*=.03), undergoing surgery in a public hospital (*P*<.001), ASA classification III and IV (*P*=.001), and having chronic diseases (*P*=.02).

Table . Association of sociodemographic and medical/surgical characteristics with desire for information among patients undergoing elective surgery at 4 participating hospitals in Northern State, Sudan, in 2024-2025 (N=305).

Characteristic	Desire for information			
	No or little desire, n (%)	Average or high desire, n (%)	Chi-square (*df*)	*P* value
Sex			1.11 (1)	.29
Male	125 (72.3)	48 (27.7)		
Female	88 (66.7)	44 (33.3)		
Age (years)			4.06 (3)	.26
18 - 30	63 (73.3)	23 (26.7)		
31 - 50	65 (63.7)	37 (36.3)		
51 - 65	40 (67.8)	19 (32.2)		
>65	45 (77.6)	13 (22.4)		
Occupation			12.22 (3)	.01
Employed	24 (52.2)	22 (47.8)		
Freelancer	90 (78.9)	24 (21.1)		
Student	21 (75.0)	7 (25.0)		
Unemployed	78 (66.7)	39 (33.3)		
Marital status			0.01 (1)	.91
Married	145 (70.0)	62 (30.0)		
Single	68 (69.4)	30 (30.6)		
Educational level			8.66 (1)	.003
Some level of education	164 (66.1)	84 (33.9)		
No education	49 (85.0)	8 (14.0)		
Average monthly income (US $)			0.71 (2)	.70
<50	88 (70.4)	37 (29.6)		
50 - 250	103 (68.2)	48 (31.8)		
>250	22 (75.9)	7 (24.1)		
Family support			4.92 (1)	.03
Yes	170 (67.2)	83 (32.8)		
No	43 (82.7)	9 (17.3)		
Type of hospital			20.46 (1)	<.001
Public hospital	100 (59.2)	69 (40.8)		
Private hospital	113 (83.1)	23 (16.9)		
ASA[a] classification			11.41 (1)	.001
ASA I and II	196 (73.1)	72 (26.9)		
ASA III and IV	17 (45.9)	20 (54.1)		
Type of surgery			1.14 (2)	.57
Minor	21 (77.8)	6 (22.2)		
Intermediate	92 (67.6)	44 (32.4)		
Major	100 (70.4)	42 (29.6)		
Type of anesthesia			0.44 (2)	.80

Characteristic	Desire for information			
	No or little desire, n (%)	Average or high desire, n (%)	Chi-square (*df*)	*P* value
General	114 (71.3)	46 (28.8)		
Spinal	84 (67.7)	40 (32.3)		
Local	15 (71.4)	6 (28.6)		
Operation covered by insurance			1.79 (1)	.18
Yes	26 (63.4)	14 (34.1)		
No	187 (70.8)	59 (22.3)		
Subjective sleep quality of the night before the operation			0.22 (1)	.64
Good	156 (70.6)	65 (29.4)		
Poor	57 (67.9)	27 (32.1)		
Previous operations			0.55 (1)	.46
Yes	86 (72.3)	33 (27.7)		
No	127 (68.3)	59 (31.7)		
Chronic diseases			0.59 (1)	.02
Yes	65 (61.3)	41 (38.7)		
No	148 (74.4)	51 (25.6)		

[a]ASA: American Society of Anesthesiologists.

The results of the logistic regression analysis (Table 6) indicated that patients who had family support were 7 times more likely to develop preoperative anxiety (aOR 7.12, 95% CI 2.64 - 19.23; *P*<.001). Patients treated in public hospitals had more than 4 times the risk of preoperative anxiety compared with those treated in private hospitals (aOR 4.31, 95% CI 2.30-8.07; *P*<.001). Patients with poor subjective sleep quality the night before the operation and those classified as ASA III and IV had more than 2 times the risk of preoperative anxiety compared with their counterparts (aOR 2.85, 95% CI 1.51 - 5.38; *P*=.001; and aOR 2.36, 95% CI 1.00 - 5.54; *P*=.049, respectively).

Table . Risk assessment of factors affecting clinically significant preoperative anxiety and desire for information, using binary and multinomial logistic regression (N=305).

Scale and variable	cOR[a] (95% CI)	*P* value	aOR (95% CI)	*P* value
Preoperative anxiety				
Family support: yes (reference: no)	2.88 (1.18 - 7.03)	.02	7.12 (2.64 - 19.23)	<.001
Type of hospital: public (reference: private)	3.35 (1.85 - 6.02)	<.001	4.31 (2.30 - 8.07)	<.001
Subjective sleep quality the night before the operation: poor (reference: good)	2.17 (1.25 - 3.78)	.01	2.85 (1.51 - 5.38)	.001
ASA[b] classification: ASA III/IV (reference: ASA I/II)	2.36 (1.16 - 4.84)	.02	2.36 (1.00 - 5.54)	.049
Chronic diseases: yes (reference: no)	1.69 (0.99 - 2.88)	.05	1.14 (0.61 - 2.14)	.61
Desire for information				
Occupation (reference: student)				
Freelancer	0.80 (0.30 - 2.10)	.65	0.78 (0.29 - 2.14)	.64
Employed	2.75 (0.98 - 7.72)	.06	2.58 (0.89 - 7.49)	.08
Unemployed	1.50 (0.59 - 3.83)	.40	1.55 (0.57 - 4.19)	.39
Educational level: some level of education (reference: no education)	3.14 (1.42 - 6.93)	.01	2.48 (1.00 - 6.11)	.049
Type of hospital: public (reference: private)	3.39 (1.97 - 5.84)	<.001	3.57 (1.93 - 6.61)	<.001
Family support: yes (reference: no)	2.33 (1.09 - 5.01)	.03	4.10 (1.81 - 9.30)	.001
ASA[b] classification: ASA III/IV (reference: ASA I/II)	3.20 (1.59 - 6.45)	<.001	3.26 (1.39 - 7.64)	.007
Chronic diseases: yes (reference: no)	1.83 (1.11 - 3.03)	.02	1.30 (0.72 - 2.37)	.39

[a]cOR: crude odds ratio.

[b]ASA: American Society of Anesthesiologists.

Patients with some level of education were more than 2 times more likely to have a desire for information than patients with no education (aOR 2.48, 95% CI 1.00 - 6.11; *P*=.049). Similarly, patients with family support were more than 4 times more likely to have a desire for information than those without family support (aOR 4.10, 95% CI 1.81 - 9.30; *P*=.001). Patients treated in public hospitals had more than 3 times the desire for information compared with those treated in private hospitals (aOR 3.57, 95% CI 1.936.61; *P*<.001). Furthermore, ASA III and IV patients were more than 3 times more likely to have a desire for information than ASA I and II patients (aOR 3.26, 95% CI 1.39 - 7.64; *P*=.007).

Discussion

Principal Findings

Preoperative anxiety is a common psychological response among patients awaiting surgery, with significant implications for their overall well-being, surgical outcomes, and recovery [22]. This study assessed the prevalence of preoperative anxiety and desire for information among patients scheduled for elective surgery in Northern State, Sudan, and identified the factors associated with these outcomes. A summary of what is known about this topic and what this study adds is shown in Textbox 1.

Textbox 1.

What Is Already Known on This Topic

- Preoperative anxiety is a common issue affecting surgical outcomes and patient well-being, with varying prevalence rates worldwide.
- Patients often desire more information than they receive, which may impact their satisfaction and anxiety levels.
- Previous studies have identified that factors such as sex, educational level, and American Society of Anesthesiologists (ASA) classification influence preoperative anxiety.

What This Study Adds

- Public and private health care settings may have different impacts on patient anxiety and information needs due to resource availability and patient experience.
- Certain cultural factors, such as extensive family involvement, may paradoxically amplify preoperative anxiety in certain regions.
- Educated patients, patients with higher ASA classification, and patients being treated in public hospitals are more likely to seek information.

Comparison With Prior Work

The study found that 24.6% of participants experienced preoperative anxiety, a rate lower than the global prevalence and that observed in LMICs (48% and 55.7%, respectively) [6,7]. However, studies from Yemen, Nigeria, Palestine, India, Nepal, and France are consistent with our results, reporting rates of 29.8%, 24.4%, 27.1%, 21.0%, 25.8%, and 24.7%, respectively [9,23-27]. These variations may be attributed to diverse factors, such as differences in health care systems, geographical locations, study populations, and assessment tools used.

Our study identified several factors linked to preoperative anxiety, including family support, poor subjective sleep quality, surgery in a public hospital, and ASA classification III and IV. Remarkably, patients with family support were 7 times more likely to experience preoperative anxiety (aOR 7.12, 95% CI 2.64 - 19.23; *P*<.001), which contrasts with the conventional view of family support as a protective factor. The unexpected result may stem from unique family dynamics in Sudan, where extended visits from relatives and community members expressing concerns or sharing negative experiences may inadvertently heighten a patient's anxiety. This underscores the complex role of social support in preoperative anxiety. Previous studies have reported conflicting results, with some suggesting that family support reduces anxiety and others indicating no significant impact [28,29].

Consistent with previous research, our study found that patients treated in public hospitals had a significantly higher risk of preoperative anxiety compared to those in private hospitals (aOR 4.31, 95% CI 2.30 - 8.07; *P*<.001). A similar trend was observed in a study from Tanzania, which also reported higher preoperative anxiety among patients in public hospitals [30]. This disparity may be due to factors such as differences in resource availability, staffing levels, and the overall patient experience between public and private health care settings.

Poor sleep quality the night before surgery was significantly associated with preoperative anxiety, consistent with previous studies highlighting the strong link between sleep disturbances and anxiety disorders [31-34]. The stress and anticipation of surgery may disrupt sleep, exacerbating anxiety levels. Similarly, patients classified as ASA III and IV had a higher risk of preoperative anxiety, likely due to their increased risk of complications and poorer health status [35]. This finding aligns with several studies demonstrating a significant association between higher ASA classification and elevated anxiety levels [23,36,37].

Notably, 23.9% of the patients expressed an average desire for information, and 6.2% reported a high desire for information. This desire, as part of broader health information–seeking behavior, enhances health literacy. However, surgeons often underestimate the extent of information patients wish to receive preoperatively [38,39]. Similar patterns have been observed in studies from Yemen and Nigeria, where comparable percentages of patients expressed average and high desire for information (16.5% and 16.9% in Yemen, and 14.4% and 5.1% in Nigeria, respectively) [9,23].

Our study reveals that patients with some education are more likely to seek information. One possible explanation is that more educated patients may feel more confident in asking questions and navigating medical discussions, whereas those with less formal education may feel less empowered or perceive themselves as less able to understand complex medical explanations. Patients in public hospitals also showed a higher desire for information, possibly due to perceived communication gaps. Additionally, ASA III and IV patients had a stronger desire for information, likely driven by the need to understand their health risks. To the best of our knowledge, this study is the first to comprehensively explore these factors, addressing a significant gap in the literature and offering insights for enhancing patient-centered care.

Conclusion

This study found that a notable proportion of surgical patients in Northern State, Sudan, experienced preoperative anxiety, a rate lower than the global average but consistent with some local and international findings. Factors associated with increased anxiety included family support, poor sleep quality, public hospital settings, and higher ASA classification. The study also highlighted a significant desire for information among patients, particularly those with some education and those receiving treatment in public hospitals. Addressing these factors through targeted interventions may improve patient outcomes and satisfaction.

Recommendations

Several recommendations can be made on the basis of these study findings to enhance patient care and reduce preoperative anxiety. Health care providers should implement targeted interventions aimed at improving preoperative sleep quality and providing adequate information to patients, as these factors were significantly linked to anxiety levels. Additionally, while family visits may help reduce preoperative anxiety for some patients, they may contribute to increased anxiety in certain situations. Clinical teams should assess the potential risks and benefits of family visits in the preoperative period and act accordingly. Furthermore, public hospitals should prioritize improving patient communication and addressing perceived gaps in information provision, as these factors were associated with a higher desire for information among patients.

Limitations and Strengths

This study has several limitations. The small sample size and the use of convenience sampling may limit generalizability, but the multicenter study design enhances representativeness. The cross-sectional design prevents establishing causality; however, it provides a detailed snapshot for future longitudinal studies. The reliance on self-reported data introduces potential biases, but using a validated assessment tool and face-to-face interviews helped minimize these issues.

Acknowledgments

The authors appreciate the participants for their valuable contributions and thank the hospital administrations for their cooperation and support in providing necessary approvals and consent. All authors declared that they had insufficient funding to support open access publication of this manuscript, including from affiliated organizations or institutions, funding agencies, or other organizations. JMIR Publications provided article processing fee (APF) support for the publication of this article.

Authors' Contributions

Conceptualization: AA (equal), MI (equal)
Formal analysis: AA (equal), MN (equal), MI (lead)
Investigation: MA (equal), AY (equal), EM (equal), AF (equal)
Methodology: AA (lead), MN (equal), AY (supporting), MI (equal)
Project administration: AA (lead), MI (supporting)
Supervision: AA (equal), MM (lead)
Writing – original draft: AA (lead), EM (supporting), AF (supporting), MM (supporting)
Writing – review & editing: MI (lead), MN (equal), MA (supporting), AY (supporting)

Conflicts of Interest

None declared.

Multimedia Appendix 1
English version of the questionnaire.
[DOCX File, 51 KB - periop_v8i1e75736_app1.docx]

References

1. Szuhany KL, Simon NM. Anxiety disorders: a review. JAMA 2022 Dec 27;328(24):2431-2445. [doi: 10.1001/jama.2022.22744] [Medline: 36573969]
2. Bandelow B, Michaelis S. Epidemiology of anxiety disorders in the 21st century. Dialogues Clin Neurosci 2015 Sep;17(3):327-335. [doi: 10.31887/DCNS.2015.17.3/bbandelow] [Medline: 26487813]
3. Friedrich S, Reis S, Meybohm P, Kranke P. Preoperative anxiety. Curr Opin Anaesthesiol 2022 Dec 1;35(6):674-678. [doi: 10.1097/ACO.0000000000001186] [Medline: 36131642]
4. Types of surgery. Royal College of Surgeons of England. URL: https://www.rcseng.ac.uk/patient-care/having-surgery/types-of-surgery/ [accessed 2025-03-04]
5. Ji W, Sang C, Zhang X, Zhu K, Bo L. Personality, preoperative anxiety, and postoperative outcomes: a review. Int J Environ Res Public Health 2022 Sep 26;19(19):12162. [doi: 10.3390/ijerph191912162] [Medline: 36231463]
6. Kil HK, Kim WO, Chung WY, Kim GH, Seo H, Hong JY. Preoperative anxiety and pain sensitivity are independent predictors of propofol and sevoflurane requirements in general anaesthesia. Br J Anaesth 2012 Jan;108(1):119-125. [doi: 10.1093/bja/aer305] [Medline: 22084330]
7. Tadesse M, Ahmed S, Regassa T, Girma T, Mohammed A. The hemodynamic impacts of preoperative anxiety among patients undergoing elective surgery: an institution-based prospective cohort study. Int J Surg Open 2022;43:100490. [doi: 10.1016/j.ijso.2022.100490]
8. Baagil H, Baagil H, Gerbershagen MU. Preoperative anxiety impact on anesthetic and analgesic use. Medicina (Kaunas) 2023 Nov 23;59(12):2069. [doi: 10.3390/medicina59122069] [Medline: 38138172]

9. Maranets I, Kain ZN. Preoperative anxiety and intraoperative anesthetic requirements. Anesth Analg 1999;89(6):1346. [doi: 10.1213/00000539-199912000-00003]
10. Chen YYK, Soens MA, Kovacheva VP. Less stress, better success: a scoping review on the effects of anxiety on anesthetic and analgesic consumption. J Anesth 2022 Aug;36(4):532-553. [doi: 10.1007/s00540-022-03081-4] [Medline: 35779126]
11. Abate SM, Chekol YA, Basu B. Global prevalence and determinants of preoperative anxiety among surgical patients: a systematic review and meta-analysis. Int J Surg Open 2020;25:6-16. [doi: 10.1016/j.ijso.2020.05.010]
12. Bedaso A, Mekonnen N, Duko B. Prevalence and factors associated with preoperative anxiety among patients undergoing surgery in low-income and middle-income countries: a systematic review and meta-analysis. BMJ Open 2022 Mar 11;12(3):e058187. [doi: 10.1136/bmjopen-2021-058187] [Medline: 35277412]
13. Lami M, Negash A, Dereje J, et al. Prevalence of preoperative anxiety and associated factors among surgical patients: systematic review and meta-analysis in Ethiopia. Health Serv Insights 2025;18:11786329251316748. [doi: 10.1177/11786329251316748] [Medline: 39906885]
14. Alqalah TAH. Prevalence, predictors, and nonpharmacological interventions for preoperative anxiety in elective noncardiac surgery patients in Yemen: path analysis. J Ment Health Human Be 2024;29(2):85-95. [doi: 10.4103/jmhhb.jmhhb_150_24]
15. von Elm E, Altman DG, Egger M, et al. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. Int J Surg 2014 Dec;12(12):1495-1499. [doi: 10.1016/j.ijsu.2014.07.013] [Medline: 25046131]
16. Cochran WG. Sampling Techniques, 3rd edition: Nashville, TN: John Wiley & Sons; 1977. URL: https://books.google.com/books/about/Sampling_Techniques.html?id=8Y4QAQAAIAAJ [accessed 2024-02-05]
17. ALMesned S, Alsalhi AA, Abdelsalam S, et al. Arabic validation of the Amsterdam Preoperative Anxiety and Information Scale. Cureus 2022 Aug;14(8):e28004. [doi: 10.7759/cureus.28004] [Medline: 36134100]
18. Kilic S. Cronbach's alpha reliability coefficient. J Mood Disord 2016;6(1):47. [doi: 10.5455/jmood.20160307122823]
19. Spanner S, Sayer L. Is the Amsterdam Preoperative Anxiety and Information Scale (APAIS) a valid tool in guiding the management of preoperative anxiety in adult patients? A literature review. J Nurs Pract 2019;3(1):95-102. [doi: 10.36959/545/368]
20. Boker A, Brownell L, Donen N. The Amsterdam preoperative anxiety and information scale provides a simple and reliable measure of preoperative anxiety. Can J Anaesth 2002 Oct;49(8):792-798. [doi: 10.1007/BF03017410] [Medline: 12374706]
21. Moerman N, van Dam FS, Muller MJ, Oosting H. The Amsterdam Preoperative Anxiety and Information Scale (APAIS). Anesth Analg 1996 Mar;82(3):445-451. [doi: 10.1097/00000539-199603000-00002] [Medline: 8623940]
22. Kassahun WT, Mehdorn M, Wagner TC, Babel J, Danker H, Gockel I. The effect of preoperative patient-reported anxiety on morbidity and mortality outcomes in patients undergoing major general surgery. Sci Rep 2022 Apr 15;12(1):6312. [doi: 10.1038/s41598-022-10302-z] [Medline: 35428818]
23. Ossai EN, Nwosu ADG, Onwuasoigwe O, Ubboe K, Ameh J, Alu L. Prevalence and predictors of anxiety among surgical patients in the preoperative holding area of National Orthopaedic Hospital, Enugu, Nigeria: a cross-sectional study. J West Afr Coll Surg 2023;13(2):105-112. [doi: 10.4103/jwas.jwas_10_23] [Medline: 37228877]
24. Shawahna R, Jaber M, Maqboul I, et al. Prevalence of preoperative anxiety among hospitalized patients in a developing country: a study of associated factors. Perioper Med (Lond) 2023 Aug 24;12(1):47. [doi: 10.1186/s13741-023-00336-w] [Medline: 37620871]
25. Omkaram S, Reddy CG, Murthy PS, Chaudhury S. Prevalence of preoperative anxiety in patients posted for surgical procedures and its relation to the doses of anesthetic drugs: a cross-sectional study. Ind Psychiatry J 2023;32(2):260-265. [doi: 10.4103/ipj.ipj_109_22] [Medline: 38161448]
26. Adhikari SP, Pathak BD, Ghimire B, et al. Prevalence of pre-operative anxiety and associated risk factors among patients awaiting elective surgery in a tertiary care hospital. F1000Res 2023;12:1207. [doi: 10.12688/f1000research.136320.2] [Medline: 38318155]
27. Dziadzko M, Mazard T, Bonhomme M, et al. Preoperative anxiety in the surgical transfer and waiting area: a cross-sectional mixed method study. J Clin Med 2022 May 9;11(9):2668. [doi: 10.3390/jcm11092668] [Medline: 35566793]
28. Rafiq M, Safdar S. Mediating role of perceived social support on mental health problems in pre-operative patients. Anaesth Pain Intensive Care 2021;25(1):65-72. [doi: 10.35975/apic.v25i1.1442]
29. Mustamu AC, Nobel Bistara D, Susanti. Family support and social support in pre-operative anxiety status. J Health Sci 2023;16(2):163-172. [doi: 10.33086/jhs.v16i02.3802]
30. Msoma RG, Khamis RH, Rwanyuma LP, Akoko L. Pre-operative anxiety predictors and hemodynamic changes at Muhimbili National Hospital. J Med Psychol 2023;12(2):55-70. [doi: 10.4236/ojmp.2023.122003]
31. Staner L. Sleep and anxiety disorders. Dialogues Clin Neurosci 2003 Sep;5(3):249-258. [doi: 10.31887/DCNS.2003.5.3/lstaner] [Medline: 22033804]
32. Kara N. Effect of sleep quality on psychiatric symptoms and life quality in newspaper couriers. Noro Psikiyatr Ars 2016 Jun;53(2):102-107. [doi: 10.5152/npa.2015.10164] [Medline: 28360780]
33. Rodrigues L, Ajith JM, Sequeira JE, Dsouza JJ, Lobo L. Pre-operative anxiety and sleep quality among the patients admitted in a selected hospital. Int J Adv Res Nurs 2024 Jul 1;7(2):205-209. [doi: 10.33545/nursing.2024.v7.i2.C.420]

34. Gu X, Zhang Y, Wei W, Zhu J. Effects of preoperative anxiety on postoperative outcomes and sleep quality in patients undergoing laparoscopic gynecological surgery. J Clin Med 2023 Feb 24;12(5):1835. [doi: 10.3390/jcm12051835] [Medline: 36902622]
35. Daabiss M. American Society of Anaesthesiologists physical status classification. Indian J Anaesth 2011 Mar;55(2):111-115. [doi: 10.4103/0019-5049.79879] [Medline: 21712864]
36. Sencaj JF, Siddique MA, Snigur GA, Ward SO, Patel SN, Singh K. Baseline American Society of Anesthesiologists classification predicts worse anxiety and pain interference following lumbar interbody fusion. J Clin Neurosci 2025 Jan;131:110929. [doi: 10.1016/j.jocn.2024.110929] [Medline: 39579392]
37. Acharya S, Gurung R, Parajuli B. Preoperative anxiety assessment in adult patients undergoing elective surgeries: a cross-sectional observational study. J Inst Med Nepal 2020;42(3):18-22. [doi: 10.59779/jiomnepal.1126]
38. Nasiri E, Birami M, Mahdavinoor SMM, Rafiei MH. Health care team understanding of patients' desire for information on surgery and anesthesia: a cross-sectional study. Perioper Care Oper Room Manag 2020 Dec;21:100134. [doi: 10.1016/j.pcorm.2020.100134]
39. Keulers BJ, Scheltinga MRM, Houterman S, Van Der Wilt GJ, Spauwen PHM. Surgeons underestimate their patients' desire for preoperative information. World J Surg 2008 Jun;32(6):964-970. [doi: 10.1007/s00268-008-9581-1] [Medline: 18408963]

Abbreviations

aOR: adjusted odds ratio
APAIS: Amsterdam Preoperative Anxiety and Information Scale
ASA: American Society of Anesthesiologists
LMICs: low- and middle-income countries
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by N Rohatgi; submitted 09.04.25; peer-reviewed by G Lakhe, M Pritchard, MC Duong; revised version received 10.08.25; accepted 25.08.25; published 15.10.25.

Please cite as:
Ahmed A, Nasur M, Mohamed E, Faragalla A, Ahmed M, Mohammed M, Yusuf A, Issak M
Preoperative Anxiety and Information Desire Among Patients Undergoing Elective Surgery in Northern Sudan: Multicenter Cross-Sectional Study
JMIR Perioper Med 2025;8:e75736
URL: https://periop.jmir.org/2025/1/e75736
doi:10.2196/75736

Original Paper

Implementation of Brief Submaximal Cardiopulmonary Testing in a High-Volume Presurgical Evaluation Clinic: Feasibility Cohort Study

Zyad James Carr[1,2], MD; Daniel Agarkov[1,2*], MD; Judy Li[2], BSc; Jean Charchaflieh[1,2*], MD; Andres Brenes-Bastos[1,2], MD; Jonah Freund[3*], BSc; Jill Zafar[1*], MD; Robert B Schonberger[1*], MD, MHCDS; Paul Heerdt[1*], MD, PhD

[1]Department of Anesthesiology, Yale University, New Haven, CT, United States

[2]School of Medicine, Yale University, New Haven, CT, United States

[3]Syracuse University, Syracuse, NY, United States

[*]these authors contributed equally

Corresponding Author:
Zyad James Carr, MD
Department of Anesthesiology
Yale University
333 Cedar Street, TMP-3
New Haven, CT, 06510
United States
Phone: 1 2037852802
Email: zyad.carr@yale.edu

Abstract

Background: Precise functional capacity assessment is a critical component for preoperative risk stratification. Brief submaximal cardiopulmonary exercise testing (smCPET) has shown diagnostic utility in various cardiopulmonary conditions.

Objective: This study aims to determine if smCPET could be implemented in a high-volume presurgical evaluation clinic and, when compared to structured functional capacity surveys, if smCPET could better discriminate low functional capacity (≤4.6 metabolic equivalents [METs]).

Methods: After institutional approval, 43 participants presenting for noncardiac surgery who met the following inclusion criteria were enrolled: aged 60 years and older, a Revised Cardiac Risk Index of ≤2, and self-reported METs of ≥4.6 (self-endorsed ability to climb 2 flights of stairs). Subjective METs assessments, Duke Activity Status Index (DASI) surveys, and a 6-minute smCPET trial were conducted. The primary end points were (1) operational efficiency, based on the time of the experimental session being ≤20 minutes; (2) modified Borg survey of perceived exertion, with a score of ≤7 indicating no more than moderate exertion; (3) high participant satisfaction with smCPET task execution, represented as a score of ≥8 (out of 10); and (4) high participant satisfaction with smCPET scheduling, represented as a score of ≥8 (out of 10). Student's *t* test was used to determine the significance of the secondary end points. Correlation between comparable structured surveys and smCPET measurements was assessed using the Pearson correlation coefficient. A Bland-Altman analysis was used to assess agreement between the methods.

Results: The mean session time was 16.9 (SD 6.8) minutes. The mean posttest modified Borg survey score was 5.35 (SD 1.8). The median patient satisfaction (on a scale of 1=worst to 10=best) was 10 (IQR 10-10) for scheduling and 10 (IQR 9-10) for task execution. Subjective METs were higher when compared to smCPET equivalents (extrapolated peak METs; mean 7.6, SD 2.0 vs mean 6.7, SD 1.8; t_{42}=2.1; P<.001). DASI-estimated peak METs were higher when compared to smCPET peak METs (mean 8.8, SD 1.2 vs mean 6.7, SD 1.8; t_{42}=7.2; P<.001). DASI-estimated peak oxygen uptake was higher than smCPET peak oxygen uptake (mean 30.9, SD 4.3 mL kg^{-1} min^{-1} vs mean 23.6, SD 6.5 mL kg^{-1} min^{-1}; t_{42}=7.2; P<.001).

Conclusions: Implementation of smCPET in a presurgical evaluation clinic is both patient centered and clinically feasible. Brief smCPET measures, supportive of published reports regarding low sensitivity of provider-driven or structured survey measures for low functional capacity, were lower than those from structured surveys. Future studies will analyze the prediction of perioperative complications and cost-effectiveness.

Trial Registration: ClinicalTrials.gov NCT05743673; https://clinicaltrials.gov/study/NCT05743673

(JMIR Perioper Med 2025;8:e65805) doi:10.2196/65805

KEYWORDS

preoperative evaluation; submaximal cardiopulmonary exercise test; risk stratification; perioperative medicine; anesthesiology

Introduction

Background

Assessment of functional capacity or exercise tolerance, as measured by self-reported metabolic equivalents (METs), remains a cornerstone of preoperative risk stratification. METs are defined as multiples of the basal metabolic rate (1 MET=3.5 mL kg^{-1} min^{-1}), and self-reported ability to climb 1 flight of stairs has a general consensus of 4 METs [1]. A threshold of ≤4.6 METs (self-reported inability to climb 2 flights of stairs) has been associated with major adverse cardiac events, all-cause mortality, and increased perioperative complications [2-4]. Despite its importance, published reports have cast doubt on the accuracy of provider-driven and self-reported assessment of functional capacity [5,6]. Thus, reliable and efficient methods to precisely characterize functional capacity continue to be of importance in preoperative risk stratification.

Cardiopulmonary exercise testing (CPET) precisely characterizes exercise tolerance by analyzing cellular respiration at rest and during exercise challenges. By measuring resting gas exchange followed by maximal exercise to expose pathophysiological impairments, CPET exploits a symptom-limited approach with a 3-minute resting stage, 3 minutes of unloaded cycling, and a 10- to 12-minute ramp stage with increasing resistance until terminated by the participant [7]. Abnormal CPET measures have been frequently associated with perioperative morbidity, with a peak oxygen uptake (VO_2) of <15 mL kg^{-1} min^{-1} reported as a threshold for elevated cardiopulmonary risk after thoracic and major noncardiac surgery [8-12]. In addition, peak VO_2 impairment predicts an increased risk of surgical site infection, postoperative respiratory failure, and critical care readmission [13]. However, CPET has not been widely adopted in preoperative testing, likely due to limited availability, required technical skills, necessity of maximal patient effort, complexity of task, and cost. Yet, conventional preoperative care, usually comprised of subjective or structured, survey-based, clinician estimation of preoperative functional capacity, has demonstrated poor sensitivity in the identification of patients with low functional capacity (≤4 METs), when compared to CPET [13,14].

In contrast to a conventional symptom-limited approach, submaximal cardiopulmonary exercise testing (smCPET) uses a time-limited approach and predictive analytics to provide estimates of peak cardiopulmonary performance [7]. A maximal exercise effort is not required since it analyzes the VO_2 efficiency slope to predict peak cardiopulmonary performance [15-17]. Of note, the VO_2 efficiency slope has a strong correlation with peak VO_2 (r=0.941), permitting effort-independent prediction of conventional CPET measures [16]. Brief smCPET has demonstrated diagnostic utility in predicting postoperative length of stay, complications, and prognosis in heart failure, pulmonary hypertension, and other conditions [18-23].

Objectives

These advantages suggest that time-limited smCPET may be useful for rapid preoperative assessment of exercise tolerance. Therefore, the primary objective was to determine the logistic feasibility of smCPET integration within a high-volume presurgical evaluation clinic. Our measured feasibility end points were (1) operational efficiency, based on the experimental session length being <20 minutes; (2) modified Borg survey of perceived exertion, with a score of ≤7 indicating no more than moderate exertion; (3) high participant satisfaction with smCPET task execution, with a score of >8 (out of 10); and (4) high patient satisfaction with smCPET scheduling, with a score of >8 (of 10). Our secondary objective was to determine if comparable smCPET measures were significantly different from structured survey findings. The secondary end points were a comparison of (1) self-reported subjective METs from a survey versus smCPET equivalents (extrapolated peak METs), (2) Duke Activity Status Index (DASI) [24] estimates versus smCPET equivalents (extrapolated peak METs), and (3) estimated DASI maximal oxygen consumption (estimated peak VO_2) versus smCPET equivalents (extrapolated peak VO_2). This study hypothesized that brief smCPET would achieve two objectives: first, meet feasibility end points indicating successful implementation, and second, similar to prior published reports regarding provider-driven functional capacity assessments, identify lower peak METs and VO_2, when compared to structured surveys.

Methods

Trial Design

This is an ongoing prospective open-label clinical device study approved by the Yale University Institutional Review Board (IRB#2000033885; ClinicalTrials.gov: #NCT05743673 [25]; principal investigator: ZJC; date of registration: December 5, 2023). This clinical trial was registered prior to participant enrollment.

Study Population

Inclusion criteria for study enrollment included age of 60 years and older, a Revised Cardiac Risk Index (RCRI) [26] of ≤2, self-endorsed subjective METs of ≥4 (endorses reliably climbing 2 flights of stairs), and presenting for noncardiac surgery. The aim was to recruit 40 participants for the feThis number was estimated to be adequate to identify any study-related logistic process problems or patient-centered outcome deficiencies and to determine the operational efficiency of this novel system process. The RCRI≤2 criterion was selected given the novelty of smCPET in preoperative evaluation.

Given that participants were screened prior to surgical procedures, exclusion criteria were adapted to maintain current

standard-of-care practices in preoperative evaluation, which includes mandatory subspecialty evaluation of select cardiopulmonary conditions. Participants with recorded severe or critical heart valve disease, active exertional angina, nonambulation, gait abnormalities, end-stage renal disease, severe peripheral vascular disease, and neurological motor deficits were excluded. Additionally, non–English-speaking participants, those under legal guardianship, and participants documented to not have personal health care decision-making capacity were also excluded. After prescreening, a phone call was placed by a study team member, and eligible participants were invited for in-person written informed consent, preoperative evaluation, questionnaire assessment of METs, and a 6-minute smCPET experimental session.

Testing Environment

Testing was performed at the presurgical evaluation (PSE) clinic at Yale New Haven Hospital, which is responsible for more than 40,000 preoperative evaluations per year. On a daily basis, the PSE clinic is staffed by an anesthesiologist, 2 resident physicians, 3 certified nurse practitioners, and 6 nursing staff and contains 6 exam rooms.

Study Apparatus

The US Food and Drug Administration–approved Shape II is a compact, cardiopulmonary, breath-by-breath, exercise testing system that uses brief submaximal exercise effort (3 minutes) to generate multiple quantitative measures of actual and predicted peak cardiopulmonary performance measurements (Figure 1). Predicted peak exercise values are automatically calculated by the device using oxygen efficiency slope equations [16,17]. Furthermore, the device has been previously validated to conventional CPET measurements [27,28]. The compact design allows all the necessary equipment to be placed on a standard rolling cart and was deployed in a PSE clinic examination room (2.4 × 2.4 m). A stairstep (14-cm height) was used for the graded exercise portion. The graded exercise was performed with a device prompt (“begin exercise”), with auditory prompts at 1-minute intervals to increase step frequency if possible. A metronome is used to provide cadence. The device provides an option for either timed or symptom-limited assessment. The timed session was selected for all participants. The timed device session requires a total of 6 minutes: 2 minutes of seated baseline resting data, 3 minutes of escalating exercise using the stairstep, and 1 minute of seated recovery data to generate a variety of individual measures of cardiac and pulmonary physiological data (Multimedia Appendix 1).

Figure 1. Performance of submaximal cardiopulmonary exercise testing requires (A) 2 minutes of calibration data in the seated position with a disposal mouthpiece connected to the device, (B) 3 minutes of graded exercise using a stair step, and 1 minute of recovery data in the seated position. The submaximal cardiopulmonary device (white and blue box) is visible on the cart, attached to a laptop with calculation software. Coauthor JF gave express permission for the use of his likeness in this simulated participant session.

Data Collection

Participants received height, weight, and vital sign measurements (heart rate, blood pressure, and pulse oximetry). Informed written consent was performed, and participants were instructed on smCPET (approximately 5 minutes).

Session time was measured from the beginning of pretest METs questionnaires until the termination of the smCPET recovery phase. A session time of ≤20 minutes would indicate that 24 high-risk participants could be screened per day per machine, permitting high-volume assessment. Session components included (1) a 7-question subjective METs assessment, (2) a 12-question DASI survey, and (3) a timed smCPET (6 minutes).

The modified Borg survey of perceived exertion was performed at session termination. After study interventions, a standard preoperative evaluation was completed, and the participant was discharged. A 24-hour postexperiment survey of minor and major complications and patient satisfaction was performed by

telephone (Multimedia Appendix 2). With the exception of the patient satisfaction survey, all survey instruments were adapted from prior publications [29-31]. DASI-estimated peak METs and peak VO_2 were calculated from individual participants' DASI scores using the recommended formula.

Statistical Analysis

End points were reported as continuous variables, described as mean (SD); ordinal variables, as median (IQR and range); and categorical variables, as number (%). Secondary end points were first analyzed using the Student *t* test (2-tailed) to compare differences in comparable measurements. Agreement between structured survey findings and smCPET comparable measurements was assessed using 2 approaches. First, a Pearson correlation coefficient was calculated to evaluate the strength and direction of the linear relationship, followed by a Bland-Altman analysis to assess agreement between methods, where differences between paired measurements were plotted against their means. Mean difference (MD) and 95% limits of agreement (LOAs) were calculated. All analyses were carried out on R (version 4.1.1; R Foundation for Statistical Computing). To reduce the introduction of bias, a complete case analysis for missing data was performed, where participants with missing data were excluded from the analysis of the respective end point. Similarly, dropouts were removed from the analysis. A *P* value of <.05 was accepted for significance.

Ethical Considerations

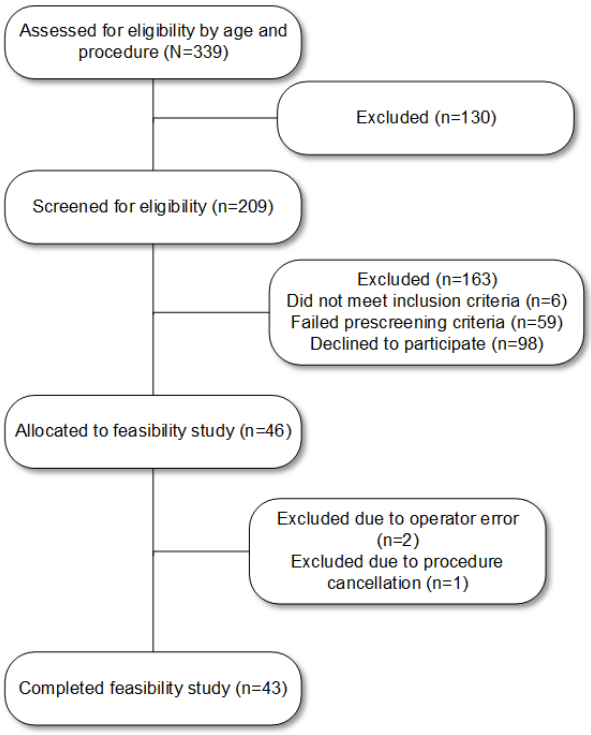
This study was performed in accordance with the principles of the Declaration of Helsinki. Approval was granted by the Yale University Institutional Review Board (IRB#2000033885). Informed consent was obtained from all participants included in the study. All provided data were deidentified prior to analysis to maintain participant privacy. No monetary compensation was provided to the participants. JF has given express written informed consent for the publication of his image in Figure 1.

Results

Participant Recruitment

We identified 209 (61.6%) out of 339 potential participants that met eligibility criteria; 6 (1.8%) did not meet the inclusion criteria, 59 (17.4%) failed the prescreening criteria, and 98 (28.9%) declined study participation (Figure 2). Initially, 46 participants were enrolled but 3 (7%) were excluded (operator error: n=2; surgery cancellation: n=1), for a final cohort of 43 participants.

Figure 2. A flow diagram of participant enrollment.



Baseline Characteristics

Trial participants had a median age of 68 (IQR 66-73, range: 60-86 years), 20 (47%) of 43 were female, and the mean BMI was 27.5 (SD 6.0) kg/m^2. Preoperative RCRI score was a median of 1 (IQR 1-1; range 1-2). Essential hypertension (22/43, 51%), hyperlipidemia (17/43, 39%), and solid tumor (25/43, 58%) were the most common premorbid conditions. A total of 22 (51%) out of 43 participants were former or active smokers. Major abdominal surgeries comprised 27 (63%) out of the 43 surgical procedures (Table 1).

All participants completed the smCPET session components. The mean peak respiratory exchange ratio was 0.88 (SD 0.12), consistent with submaximal effort (respiratory exchange ratio<1.05). The ventilatory threshold was achieved in 22 (51%) of 43 participants (mean 227.9, SD 21.9 seconds in those that achieved ventilatory threshold).

Table 1. Baseline demographical data of the study cohort (n=43).

Variable	Values
Age (years), median (IQR; range)	68 (66-73; 60-86)
Sex, n (%)	
Male	23 (54)
Female	20 (47)
BMI (kg/m^2), mean (SD)	27.5 (6.0)
Revised Cardiac Risk Index score, median (IQR; range)	1 (1-1; 1-2)
Preoperative comorbidities, n (%)	
Essential hypertension	22 (51)
Hyperlipidemia	17 (40)
Ventricular dysrhythmia	1 (2)
Congestive heart failure	1 (2)
Myocardial infarction	3 (7)
Cerebrovascular disease	1 (2)
Chronic obstructive pulmonary disease	3 (7)
Asthma	4 (9)
Obstructive sleep apnea	3 (7)
History of prior lung resection	1 (2)
Diabetes mellitus	7 (16)
Thyroid disorders	7 (16)
Solid tumor	25 (58)
Anemia	1 (2)
Social history, n (%)	
Smoking	
Ever	22 (51)
Active	4 (9)
Former	18 (42)
Never	21 (49)
Marijuana use (active)	4 (9)
Alcohol use	
Active	24 (56)
Former	16 (37)
Never	3 (7)
Cardiovascular medication use, n (%)	
Beta-blocker	14 (33)
Calcium channel antagonist	9 (21)
ACE/ARB[a] antagonist	16 (37)
Diuretic	12 (28)
Surgical categories, n (%)	
Abdominal major	27 (63)
Musculoskeletal major	4 (9)
Neurosurgical major	2 (5)

Variable	Values
Thoracic major	5 (12)
Other major	5 (12)

[a]ACE/ARB; angiotensin converting enzyme inhibitor/angiotensin receptor blockers.

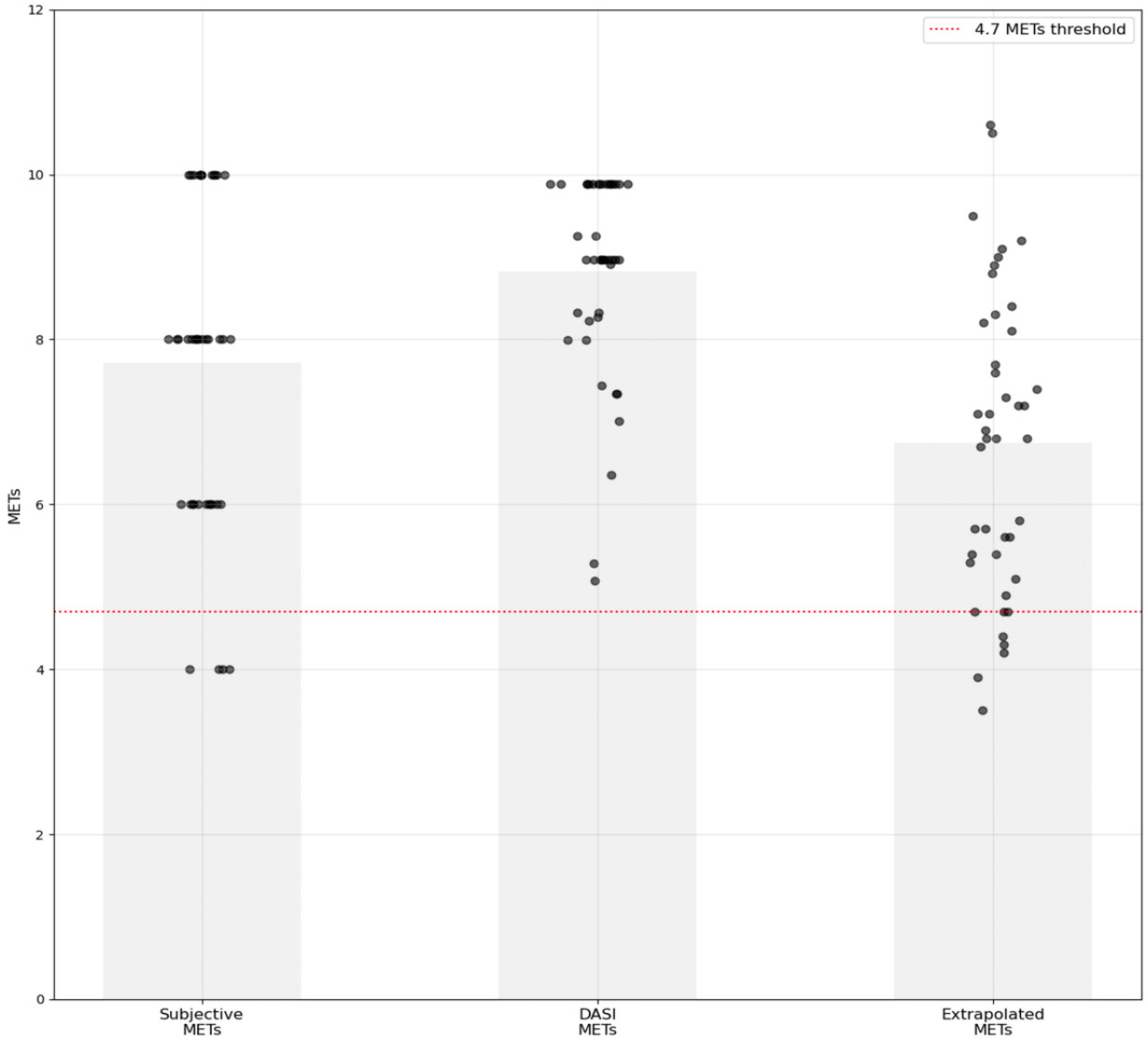
Primary End Points

The mean experimental session time was 16.9 (SD 6.8) minutes. The modified Borg survey score after experimental sessions was mean 5.35 (SD 1.8), corresponding to moderate perceived exertion. All 43 participants were reached for the 24-hour postexperiment survey. The median patient satisfaction (on a scale of 1=worst to 10=best) was 10 (IQR 10-10) for scheduling and 10 (IQR 9-10) for task execution. No major or minor complications associated with study testing were reported by participants. Operational efficiency was achieved within 15 experimental sessions among 4 study team members (3 physicians and 1 undergraduate researcher).

Secondary End Points

Average self-reported peak METs were higher when compared to smCPET equivalents (extrapolated peak METs; mean 7.6, SD 2.0 vs mean 6.7, SD 1.8; t_{42}=2.1; P<.001). DASI-estimated peak METs were higher when compared to the smCPET equivalents (extrapolated peak METs; mean 8.8, SD 1.2 vs mean 6.7, SD 1.8; t_{42}=7.2; P<.001). DASI-estimated peak VO_2 was higher than the smCPET equivalent (extrapolated peak VO_2; mean 30.9, SD 4.3 mL kg^{-1} min^{-1} vs mean 23.6, SD 6.5 mL kg^{-1} min^{-1}; t_{42}=2.1; P<.001). Figure 3 provides a comparison of values obtained from smCPET compared to structured survey–estimated peak METs and DASI-estimated peak METs.

Figure 3. Comparison of elicited METs from 2 structured survey instruments (subjective METs and DASI) compared to predicted peak METs from submaximal cardiopulmonary exercise testing (dotted line represents 4.7 METs). DASI: Duke Activity Status Index; MET: metabolic equivalent.



To analyze the congruency between the 3 study instruments, correlation and Bland-Altman analyses were performed. DASI-estimated METs showed a moderate positive correlation versus subjective METs (r=0.63; P<.001). Weaker correlations were observed with smCPET-derived extrapolated peak METs versus DASI and subjective METs (r=0.29; P=.06 and r=0.144; P=.36, respectively). DASI versus subjective METs showed an MD of 1.1 (SD 1.49; 95% LOAs –1.82 to 4.02) METs, while DASI versus smCPET-derived extrapolated peak METs showed larger discrepancies with an MD of 2.07 (SD 1.86; 95% LOAs –1.58 to 5.73) METs. The comparison between subjective METs and smCPET-derived extrapolated peak METs showed intermediate systematic bias with the widest LOAs (MD 0.97, SD 2.43 METs; 95% LOAs –3.80 to 5.75). When comparing DASI and smCPET-derived extrapolated peak VO_2 values, a positive MD was observed, indicating that DASI estimates were consistently higher (MD 7.23, SD 6.54 mL kg^{-1} min^{-1}; 95% LOA –8.11 to 21.12) and showed poor agreement (r=0.28; Multimedia Appendix 3).

Discussion

Principal Findings

Integration of brief smCPET in a high-volume PSE clinic was feasible as measured by the primary end points of session time, patient satisfaction with smCPET task execution, perceived exertion, and session scheduling. The operational efficiency of study team members was acceptable within 15 experimental sessions. Finally, smCPET measures of peak METs and VO_2 were significantly lower, when compared to comparable structured survey results.

Mean session time, which included the subjective METs survey, DASI, and 6-minute smCPET session, was 16.9 (SD 6.8) minutes, with progressive improvement over the study time period as operators (n=4) became facile with the study instrument (Multimedia Appendix 4). It is important to note that smCPET comprised 6 minutes of the session time, shorter than reported times with conventional CPET (15-20 min/session) [32]. In high-volume PSE, this may be advantageous, as patients are often seen on short notice for preoperative evaluation. Participants were able to flexibly arrange smCPET around other clinic appointments, decreasing study participants' time constraints. This likely enhanced our high satisfaction score for scheduling. High patient satisfaction was observed with task execution and perceived exertion during smCPET. The tested device uses a stationary stairstep for graded exercise, which was frequently familiar to participants. The short duration of graded exercise (3 minutes) was not perceived by any participant as maximum exertion by the Borg survey, likely contributing to the high level of patient satisfaction. Second, the Borg score of <7 after smCPET suggests a reasonable probability of success when transitioning its use to patients with more severe comorbidities, or preoperative deconditioning. It is important to note that the ventilatory threshold, or anaerobic threshold, was not measurable in 50% of our cohort, suggesting that the brief graded exercise contributed to the reported exertion level and high participant satisfaction.

One of the goals of smCPET is to make precise cardiopulmonary evaluation more widely available and patient centered, advantages that are acknowledged by its increasing adoption in the routine assessment of heart failure and pulmonary hypertension. Consistent with large-scale CPET application in cardiovascular clinical trials, smCPET did not result in findings of major or minor complications despite encouraging participants to safely provide their best effort within the timed and graded exercise component [33]. This is reassuring, as early termination of preoperative CPET trials, due to participant fatigue, safety, or other considerations, has been reported to be approximately 11% [13]. However, we purposefully selected functionally independent participants with self-reported ≥4.6 METs, and expansion to patients who are less functionally independent may result in higher smCPET session failure rates. Regardless, the safety of smCPET has been suggested by its routine application to high-risk and frail populations with severe cardiopulmonary disease, suggesting that a wide spectrum of preoperative populations can be safely tested using smCPET [20,22,34].

The structured survey estimated METs were, on average, significantly higher than their smCPET equivalents. Using the subjective METs structured survey, 8 (19%) of 43 participants reported peak METs within 10% of smCPET extrapolated peak METs, 12 (28%) were underestimated by >10%, and 23 (53%) were overestimated by >10%, when compared to smCPET values. Brief smCPET identified that 8 (19%) out of 43 study participants had ≤4.6 extrapolated peak METs (peak VO_2 equivalent: 14 mL kg^{-1} min^{-1}), corresponding to a METs threshold associated with higher perioperative cardiovascular risk [1,4]. Furthermore, smCPET identified 9 (21%) out of 43 participants with an age-adjusted peak VO_2 of less than 20 mL kg^{-1} min^{-1}, corresponding to poor aerobic capacity, and 2 (5%) with an extrapolated peak VO_2 less than 15 mL kg^{-1} min^{-1}, a measure frequently associated with higher perioperative risk [35]. These findings support prior descriptions of provider-driven and structured survey overestimation bias, highlighting the challenge of obtaining an accurate preoperative functional capacity assessment. Clinicians, when compared to CPET, had a 19.2% sensitivity in identifying low functional capacity (≤4 METs) [13,36]. Other investigations have also observed that preanesthesia evaluation calculation of self-reported METs overestimate functional capacity when compared to CPET assessment [6]. DASI was also found to poorly predict participants with lower peak VO_2 [13,24,36]. In a cohort of participants that would not necessarily receive extensive preoperative assessment, given that 100% reported the ability to reliably climb 2 flights of stairs, this may suggest opportunities to identify and preemptively optimize unexpected cardiopulmonary impairments prior to surgical intervention.

Worldwide, value-based health care has been a significant priority, and conventional preoperative evaluation may increase overall testing costs without improving perioperative outcomes [37-39]. Implementing brief smCPET for individualized preoperative cardiovascular evaluation may improve the precision of preoperative cardiovascular risk assessment and may potentially curb excess preoperative cardiovascular testing

commonly associated with older age and patients with higher comorbidities [40-42]. However widespread adoption of this technology in the perioperative space will require (1) further evidence of smCPET predictive validity for perioperative outcomes, (2) characterization of optimal system processes for patient selection, and (3) justification of cost-benefit.

Study Limitations

Several study limitations limit generalizability to other populations. Selection bias should be acknowledged given that participants who volunteered for the study are likely to be more health-conscious than usual patients who undergo PSE. A measurement bias may be introduced into the study given that researchers may unconsciously influence participant performance on smCPET or interpret results differently based on unconscious expectations. Similarly, a recall bias is often introduced when using structured, interview-style questionnaires such as those used in our study. Instrument bias may similarly impact smCPET findings; however, this is substantially reduced by routine device calibration.

Confounding factors are similar, where participants with higher fitness levels would find it easier to adapt to the stairstep exercise challenge. Our inclusion criteria purposely selected participants with lower comorbidities to ensure successful participation rates for this feasibility study. We acknowledge that certain premorbid conditions and chronic medication usage may influence smCPET participants' performance, but we did not balance this factor in this exploratory study. Although CPET and smCPET predictive performance with cardiovascular perioperative morbidity and mortality has been previously published, our cohort is not yet powered for the assessment of perioperative outcomes with this device [19,23,43,44]. Finally, the finding of no device-related adverse events should be cautiously interpreted given the small sample size and the possibility of rare exercise-induced adverse events.

Conclusions

In summary, we observed that smCPET implementation was well accepted into the workflow of a high-volume PSE clinic. Operator efficiency with the smCPET instrument was rapid and achieved relative parity at 15 participant sessions. smCPET, when compared to usual session times for conventional CPET of 15-20 minutes, uses less than half the time (6 minutes), making it attractive for the purposes of precise but time-efficient preoperative evaluation of exercise tolerance. This feasibility analysis has (1) reinforced the operational integrity of our active study protocol assessing smCPET findings with perioperative outcomes and (2) affirmed satisfactory patient-centered outcomes with study procedures. Studies should further expand smCPET predictive validity to postoperative cardiopulmonary complications, assess cost-effectiveness, and develop optimal system processes for patient selection.

Acknowledgments

The authors would like to acknowledge the work of Rayna Lewoc, Elizabeth Womack, Elizabeth Cozza, Gwendolyn Burkett, and the staff of the Yale New Haven Hospital's presurgical evaluation clinic. This study was partially supported by Shape Medical Systems, Inc.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Study conception and design were performed by ZJC, PH, and RBS. The first draft of the manuscript was performed by ZJC, JC, JZ, and ABB. Data collection was performed by ZJC, JF, JL, and DA. Statistical analysis was performed by ZJC. Further manuscript editing was performed by all coauthors. All authors read and approved the final manuscript.

Conflicts of Interest

ZJC receives partial funding from Shape Medical Systems, Inc related to this work. RBS reports owning stock in Johnson and Johnson unrelated to this work. RBS reports that Yale University has received funding from Merck for a study in which he was involved, unrelated to this work. PH reported receiving research support grants from Edwards Lifesciences and consulting and/or royalty fees from Baudax Bio, Fire1Foundry, Cardiage LLC, and Edwards Lifesciences. All other authors have no competing interests.

Multimedia Appendix 1
A detailed description of the submaximal cardiopulmonary exercise testing device and its measures.
[DOCX File , 18 KB - periop_v8i1e65805_app1.docx]

Multimedia Appendix 2
Adapted subjective METs survey questions and 24 hour postsession minor and major adverse events survey. MET: metabolic equivalent.
[PDF File (Adobe PDF File), 237 KB - periop_v8i1e65805_app2.pdf]

Multimedia Appendix 3
Bland-Altman plots for the compared measures: (A) subjective METs versus smCPET-extrapolated peak METs showed a mean difference of 0.97 METs but the widest LOAs; (B) DASI–estimated peak METs versus smCPET-extrapolated peak METs showed a mean difference of 2.07 METs, the largest discrepancy; (C) DASI-estimated peak METs versus subjective METs showed a mean difference of 1.1 METs but narrower LOAs; and (D) DASI-estimated peak VO2 versus smCPET-extrapolated peak VO2 showed that DASI had consistently higher estimates, with a mean difference of 6.5 mL kg–1 min–1. DASI: Duke Activity Status Index; LOA: limit of agreement; mean diff: mean difference; MET: metabolic equivalent; smCPET: submaximal cardiopulmonary exercise testing; VO2: peak oxygen uptake.
[PDF File (Adobe PDF File), 136 KB - periop_v8i1e65805_app3.pdf]

Multimedia Appendix 4
Operator efficiency as a function of session time (y-axis), defined as performance of two structured functional capacity survey instruments and submaximal cardiopulmonary exercise testing, versus session number (x-axis).
[PNG File , 98 KB - periop_v8i1e65805_app4.png]

References

1. Fleisher LA, Fleischmann KE, Auerbach AD, Barnason SA, Beckman JA, Bozkurt B, et al. 2014 ACC/AHA guideline on perioperative cardiovascular evaluation and management of patients undergoing noncardiac surgery: a report of the American College of Cardiology/American Heart Association Task Force on practice guidelines. J Am Coll Cardiol 2014;64(22):e77-137 [FREE Full text] [doi: 10.1016/j.jacc.2014.07.944] [Medline: 25091544]
2. Lurati Buse GAL, Puelacher C, Gualandro DM, Genini AS, Hidvegi R, Bolliger D, et al. Association between self-reported functional capacity and major adverse cardiac events in patients at elevated risk undergoing noncardiac surgery: a prospective diagnostic cohort study. Br J Anaesth 2021;126(1):102-110 [FREE Full text] [doi: 10.1016/j.bja.2020.08.041] [Medline: 33081973]
3. Reilly DF, McNeely MJ, Doerner D, Greenberg DL, Staiger TO, Geist MJ, et al. Self-reported exercise tolerance and the risk of serious perioperative complications. Arch Intern Med 1999;159(18):2185-2192. [doi: 10.1001/archinte.159.18.2185] [Medline: 10527296]
4. Biccard BM. Relationship between the inability to climb two flights of stairs and outcome after major non-cardiac surgery: implications for the pre-operative assessment of functional capacity. Anaesthesia 2005;60(6):588-593 [FREE Full text] [doi: 10.1111/j.1365-2044.2005.04181.x] [Medline: 15918830]
5. Weinstein AS, Sigurdsson MI, Bader AM. Comparison of preoperative assessment of patient's metabolic equivalents (METs) estimated from history versus measured by exercise cardiac stress testing. Anesthesiol Res Pract 2018;2018:5912726 [FREE Full text] [doi: 10.1155/2018/5912726] [Medline: 30250484]
6. Nieves-Alonso JM, Méndez Hernández RM, Ramasco Rueda F, Planas Roca A. Estimated metabolic equivalents of task do not correlate with the maximal oxygen consumption of patients undergoing lung resection surgery. Rev Esp Anestesiol Reanim 2022;69(7):437-441. [doi: 10.1016/j.redare.2021.01.006] [Medline: 35869005]
7. Glaab T, Taube C. Practical guide to cardiopulmonary exercise testing in adults. Respir Res 2022;23(1):9 [FREE Full text] [doi: 10.1186/s12931-021-01895-6] [Medline: 35022059]
8. Loewen GM, Watson D, Kohman L, Herndon JE, Shennib H, Kernstine K, et al. Preoperative exercise Vo2 measurement for lung resection candidates: results of Cancer and Leukemia Group B Protocol 9238. J Thorac Oncol 2007;2(7):619-625 [FREE Full text] [doi: 10.1097/JTO.0b013e318074bba7] [Medline: 17607117]
9. Benzo R, Kelley GA, Recchi L, Hofman A, Sciurba F. Complications of lung resection and exercise capacity: a meta-analysis. Respir Med 2007;101(8):1790-1797 [FREE Full text] [doi: 10.1016/j.rmed.2007.02.012] [Medline: 17408941]
10. Hartley RA, Pichel AC, Grant SW, Hickey GL, Lancaster PS, Wisely NA, et al. Preoperative cardiopulmonary exercise testing and risk of early mortality following abdominal aortic aneurysm repair. Br J Surg 2012;99(11):1539-1546. [doi: 10.1002/bjs.8896] [Medline: 23001820]
11. Older P, Hall A, Hader R. Cardiopulmonary exercise testing as a screening test for perioperative management of major surgery in the elderly. Chest 1999;116(2):355-362. [doi: 10.1378/chest.116.2.355] [Medline: 10453862]
12. Argillander TE, Heil TC, Melis RJF, van Duijvendijk P, Klaase JM, van Munster BC. Preoperative physical performance as predictor of postoperative outcomes in patients aged 65 and older scheduled for major abdominal cancer surgery: a systematic review. Eur J Surg Oncol 2022;48(3):570-581 [FREE Full text] [doi: 10.1016/j.ejso.2021.09.019] [Medline: 34629224]
13. Wijeysundera DN, Pearse RM, Shulman MA, Abbott TEF, Torres E, Ambosta A, et al. Assessment of functional capacity before major non-cardiac surgery: an international, prospective cohort study. Lancet 2018;391(10140):2631-2640. [doi: 10.1016/S0140-6736(18)31131-0] [Medline: 30070222]
14. Struthers R, Erasmus P, Holmes K, Warman P, Collingwood A, Sneyd JR. Assessing fitness for surgery: a comparison of questionnaire, incremental shuttle walk, and cardiopulmonary exercise testing in general surgical patients. Br J Anaesth 2008;101(6):774-780 [FREE Full text] [doi: 10.1093/bja/aen310] [Medline: 18953057]

15. Ferguson M, Shulman M. Cardiopulmonary exercise testing and other tests of functional capacity. Curr Anesthesiol Rep 2022;12(1):26-33 [FREE Full text] [doi: 10.1007/s40140-021-00499-6] [Medline: 34840532]
16. Baba R, Nagashima M, Goto M, Nagano Y, Yokota M, Tauchi N, et al. Oxygen uptake efficiency slope: a new index of cardiorespiratory functional reserve derived from the relation between oxygen uptake and minute ventilation during incremental exercise. J Am Coll Cardiol 1996;28(6):1567-1572 [FREE Full text] [doi: 10.1016/s0735-1097(96)00412-3] [Medline: 8917273]
17. Hollenberg M, Tager IB. Oxygen uptake efficiency slope: an index of exercise performance and cardiopulmonary reserve requiring only submaximal exercise. J Am Coll Cardiol 2000;36(1):194-201 [FREE Full text] [doi: 10.1016/s0735-1097(00)00691-4] [Medline: 10898434]
18. Snowden CP, Prentis JM, Anderson HL, Roberts DR, Randles D, Renton M, et al. Submaximal cardiopulmonary exercise testing predicts complications and hospital length of stay in patients undergoing major elective surgery. Ann Surg 2010;251(3):535-541. [doi: 10.1097/SLA.0b013e3181cf811d] [Medline: 20134313]
19. Prentis JM, Manas DMD, Trenell MI, Hudson M, Jones DJ, Snowden CP. Submaximal cardiopulmonary exercise testing predicts 90-day survival after liver transplantation. Liver Transpl 2012;18(2):152-159 [FREE Full text] [doi: 10.1002/lt.22426] [Medline: 21898768]
20. Oakland HT, Joseph P, Elassal A, Cullinan M, Heerdt PM, Singh I. Diagnostic utility of sub-maximum cardiopulmonary exercise testing in the ambulatory setting for heart failure with preserved ejection fraction. Pulm Circ 2020;10(4):2045894020972273 [FREE Full text] [doi: 10.1177/2045894020972273] [Medline: 33282205]
21. Joseph P, Savarimuthu S, Zhao J, Yan X, Oakland HT, Cullinan M, et al. Noninvasive determinants of pulmonary hypertension in interstitial lung disease. Pulm Circ 2023;13(1):e12197 [FREE Full text] [doi: 10.1002/pul2.12197] [Medline: 36814586]
22. Bernstein EJ, Mandl LA, Gordon JK, Spiera RF, Horn EM. Submaximal heart and pulmonary evaluation: a novel noninvasive test to identify pulmonary hypertension in patients with systemic sclerosis. Arthritis Care Res 2013;65(10):1713-1718 [FREE Full text] [doi: 10.1002/acr.22051] [Medline: 23740875]
23. Woods PR, Bailey KR, Wood CM, Johnson BD. Submaximal exercise gas exchange is an important prognostic tool to predict adverse outcomes in heart failure. Eur J Heart Fail 2011;13(3):303-310 [FREE Full text] [doi: 10.1093/eurjhf/hfq187] [Medline: 21036777]
24. Hlatky MA, Boineau RE, Higginbotham MB, Lee KL, Mark DB, Califf RM, et al. A brief self-administered questionnaire to determine functional capacity (the Duke Activity Status Index). Am J Cardiol 1989;64(10):651-654. [doi: 10.1016/0002-9149(89)90496-7] [Medline: 2782256]
25. A feasibility study of the SHAPE™ test of aerobic fitness in older adults presenting for moderate to high-risk surgery. ClinicalTrials.gov. 2023. URL: https://clinicaltrials.gov/study/NCT05743673 [accessed 2025-01-21]
26. Ford MK, Beattie WS, Wijeysundera DN. Systematic review: prediction of perioperative cardiac complications and mortality by the revised cardiac risk index. Ann Intern Med 2010;152(1):26-35. [doi: 10.7326/0003-4819-152-1-201001050-00007] [Medline: 20048269]
27. Miller AD, Woods PR, Olson TP, Hulsebus ML, O'Malley KA, MacCarter D, et al. Validation of a simplified, portable cardiopulmonary gas exchange system for submaximal exercise testing. Open Sport Med J 2010;4(1):34-40. [doi: 10.2174/1874387001004010034]
28. Kim CH, Hansen JE, MacCarter DJ, Johnson BD. Algorithm for predicting disease likelihood from a submaximal exercise test. Clin Med Insights Circ Respir Pulm Med 2017;11:1179548417719248 [FREE Full text] [doi: 10.1177/1179548417719248] [Medline: 28757799]
29. Borg GA. Psychophysical bases of perceived exertion. Med Sci Sports Exerc 1982;14(5):377-381. [Medline: 7154893]
30. McAuley P, Myers J, Abella J, Froelicher V. Evaluation of a specific activity questionnaire to predict mortality in men referred for exercise testing. Am Heart J 2006;151(4):890.e1-890897. [doi: 10.1016/j.ahj.2005.09.017] [Medline: 16569555]
31. Bell EC, Cox NS, Goh N, Glaspole I, Westall GP, Watson A, et al. Oxygen therapy for interstitial lung disease: a systematic review. Eur Respir Rev 2017;26(143):160080 [FREE Full text] [doi: 10.1183/16000617.0080-2016] [Medline: 28223395]
32. Albouaini K, Egred M, Alahmar A, Wright DJ. Cardiopulmonary exercise testing and its application. Heart 2007;93(10):1285-1292 [FREE Full text] [doi: 10.1136/hrt.2007.121558] [Medline: 17890705]
33. O'Connor CM, Whellan DJ, Lee KL, Keteyian SJ, Cooper LS, Ellis SJ, et al. Efficacy and safety of exercise training in patients with chronic heart failure: HF-ACTION randomized controlled trial. JAMA 2009;301(14):1439-1450 [FREE Full text] [doi: 10.1001/jama.2009.454] [Medline: 19351941]
34. Woods PR, Frantz RP, Taylor BJ, Olson TP, Johnson BD. The usefulness of submaximal exercise gas exchange to define pulmonary arterial hypertension. J Heart Lung Transplant 2011;30(10):1133-1142 [FREE Full text] [doi: 10.1016/j.healun.2011.03.021] [Medline: 21622009]
35. Kaminsky LA, Arena R, Myers J, Peterman JE, Bonikowske AR, Harber MP, et al. Updated reference standards for cardiorespiratory fitness measured with cardiopulmonary exercise testing: data from the Fitness Registry and the Importance of Exercise National Database (FRIEND). Mayo Clin Proc 2022;97(2):285-293 [FREE Full text] [doi: 10.1016/j.mayocp.2021.08.020] [Medline: 34809986]

36. Wijeysundera DN, Beattie WS, Hillis GS, Abbott TEF, Shulman MA, Ackland GL, Measurement of Exercise Tolerance before Surgery Study Investigators, et al. Integration of the Duke Activity Status Index into preoperative risk evaluation: a multicentre prospective cohort study. Br J Anaesth 2020;124(3):261-270 [FREE Full text] [doi: 10.1016/j.bja.2019.11.025] [Medline: 31864719]
37. World Economic Forum. Value in Healthcare: Laying the Foundation for Health System Transformation. Geneva, Switzerland: World Economic Forum; 2017.
38. Mahajan A, Esper SA, Cole DJ, Fleisher LA. Anesthesiologists' role in value-based perioperative care and healthcare transformation. Anesthesiology 2021;134(4):526-540. [doi: 10.1097/ALN.0000000000003717] [Medline: 33630039]
39. Beckerleg W, Kobewka D, Wijeysundera DN, Sood MM, McIsaac DI. Association of preoperative medical consultation with reduction in adverse postoperative outcomes and use of processes of care among residents of Ontario, Canada. JAMA Intern Med 2023;183(5):470-478 [FREE Full text] [doi: 10.1001/jamainternmed.2023.0325] [Medline: 36972037]
40. Ellenbogen MI, Drmanovic A, Segal JB, Kapoor S, Wagner PC. Patient, provider, and system-level factors associated with preoperative cardiac testing: a systematic review. J Hosp Med 2023;18(11):1021-1033. [doi: 10.1002/jhm.13206] [Medline: 37728150]
41. Kumar M, Wilkinson K, Li YH, Masih R, Gandhi M, Saadat H, et al. Association of a novel electronic form for preoperative cardiac risk assessment with reduction in cardiac consultations and testing: retrospective cohort study. JMIR Perioper Med 2024;7:e63076 [FREE Full text] [doi: 10.2196/63076] [Medline: 39269754]
42. Ferré F, Laurent R, Furelau P, Doumard E, Ferrier A, Bosch L, et al. Perioperative risk assessment of patients using the MyRISK digital score completed before the preanesthetic consultation: prospective observational study. JMIR Perioper Med 2023;6:e39044 [FREE Full text] [doi: 10.2196/39044] [Medline: 36645704]
43. Bédat B, Koliakos E, Demarchi MS, Perentes J, Licker MJ, Triponez F, et al. Ventilatory efficiency slope is associated with cardiopulmonary complications after thoracoscopic anatomical lung resection. Interact Cardiovasc Thorac Surg 2022;35(1):ivac039 [FREE Full text] [doi: 10.1093/icvts/ivac039] [Medline: 35157073]
44. Kallianos A, Rapti A, Tsimpoukis S, Charpidou A, Dannos I, Kainis E, et al. Cardiopulmonary exercise testing (CPET) as preoperative test before lung resection. In Vivo 2014;28(6):1013-1020 [FREE Full text] [Medline: 25398794]

Abbreviations

CPET: cardiopulmonary exercise testing
DASI: Duke Activity Status Index
LOA: limit of agreement
MD: mean difference
MET: metabolic equivalent
PSE: presurgical evaluation
RCRI: Revised Cardiac Risk Index
smCPET: submaximal cardiopulmonary exercise testing
VO_2: oxygen uptake

Edited by T Aslanidis; submitted 26.08.24; peer-reviewed by DJ Elmer, N Agnew; comments to author 18.12.24; revised version received 21.12.24; accepted 08.01.25; published 17.02.25.

Please cite as:
Carr ZJ, Agarkov D, Li J, Charchaflieh J, Brenes-Bastos A, Freund J, Zafar J, Schonberger RB, Heerdt P
Implementation of Brief Submaximal Cardiopulmonary Testing in a High-Volume Presurgical Evaluation Clinic: Feasibility Cohort Study
JMIR Perioper Med 2025;8:e65805
URL: https://periop.jmir.org/2025/1/e65805
doi:10.2196/65805
PMID:39773953

Quantification of Metamorphopsia Using a Smartphone-Based Hyperacuity Test in Patients With Idiopathic Epiretinal Membranes: Prospective Observational Study

Daria Amon[1], BSc, Dr med; Christoph Leisser[1], Priv Doz, Dr med, PhD; Andreas Schlatter[1], Dr med; Manuel Ruiss[1], MSc; Caroline Pilwachs[1]; Natascha Bayer[1], MSc; Josef Huemer[2,3], Dr med; Oliver Findl[1], MBA, Prof Dr med

[1]Vienna Institute for Research in Ocular Surgery, A Karl Landsteiner Institute, Hanusch Hospital, Vienna, Austria

[2]NIHR Biomedical Research Center at Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, 11-43 Bath Street Lower Ground Floor, London, United Kingdom

[3]Department of Ophthalmology and Optometry, Kepler University Hospital, Linz, Austria

Corresponding Author:
Josef Huemer, Dr med
NIHR Biomedical Research Center at Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, 11-43 Bath Street Lower Ground Floor, London, United Kingdom

Abstract

Background: Quality of vision in patients with idiopathic epiretinal membranes (iERMs) is closely linked to distorted vision (metamorphopsia), which is often underestimated in clinical settings. Current surgical decision-making relies heavily on visual acuity and optical coherence tomography findings, which do not adequately reflect the patient's functional vision or the severity of metamorphopsia. There is a clinical need for tools that can reliably quantify this symptom to improve patient outcomes and streamline care pathways.

Objectives: This study is the first to assess the use of a novel smartphone-based hyperacuity test (SHT) in quantifying metamorphopsia before and after surgical intervention for iERMs, comparing it with a conventional printed chart.

Methods: This prospective observational study included 27 patients with iERMs with symptomatic metamorphopsia detected on the Amsler grid scheduled for vitrectomy with membrane peeling. The SHT (Alleye, Oculocare Medical Inc) and the horizontal (MH) and vertical (MV) M-chart (Inami & Co, Ltd) tests were performed 3 times before and 3 months after surgery. Pre- and postoperative metamorphopsia scores, changes in distance-corrected visual acuity, optical coherence tomography biomarkers, and subjective perception of metamorphopsia were evaluated.

Results: The mean SHT score significantly (r=0.686; P<.001) improved from 55.2 (SD 18.9) before surgery to 63.5 (SD 16.3) after surgery while the improvement of the M-chart scores were insignificant (MH r=0.37, P=.06; MV r=0.18, P=.36). Pre- and postoperative SHT scores showed very weak and insignificant correlations with the MH, MV, and MH+MV scores. Both metamorphopsia tests showed good reliability (intraclass correlation coefficients >0.75).

Conclusions: The SHT showed a significant improvement in postoperative metamorphopsia scores, indicating that it could be a valuable tool for quantifying visual distortion in patients with iERMs. While discrepancies with M-chart results were observed, both tests demonstrated good reliability. Clinically, the SHT may offer a practical solution for monitoring metamorphopsia and guiding complex surgical decision-making, particularly in telemedicine settings. Its accessibility could improve patient management, potentially enhancing preoperative triaging and reducing unnecessary visits.

Trial Registration: ClinicalTrials.gov NCT05138315; https://clinicaltrials.gov/study/NCT05138315

(JMIR Perioper Med 2025;8:e60959) doi:10.2196/60959

KEYWORDS

mobile health; smartphone; telemedicine; Alleye; M-chart; metamorphopsia; epiretinal membrane; vitrectomy with membrane peeling; visual acuity; home monitoring; hyperacuity test; hyperacuity; surgical intervention; distorted vision; vision; ocular pathology; ocular; retinal; retina; surgery; macular degeneration; tomography; vitrectomy; ophthalmology

Introduction

An epiretinal membrane (ERM) is a common disorder leading to a decrease in visual acuity and distorted vision, called "metamorphopsia" in later stages [1,2]. ERMs in the majority of cases are idiopathic with no identifiable cause; however, they may be secondary due to an already existing ocular pathology. Metamorphopsia, one of the leading symptoms of ERM, can

be detected with the Amsler grid or the metamorphopsia chart (M-chart; Inami & Co, Ltd); however, these tools, including the M-chart, are often not routinely used in clinical practice despite metamorphopsia's importance in surgical decision-making. Clinicians often do not routinely assess metamorphopsia despite its impact on surgical decisions, relying instead on visual acuity, which often fails to reflect the subjective and functionally significant experience of metamorphopsia. Visualization of the membrane and evaluation of microstructural retinal alterations are best achieved with optical coherence tomography (OCT) but do not fully capture the functional visual disturbances experienced by patients [3]. ERMs are surgically treated via vitrectomy and membrane peeling. The degree of postoperative metamorphopsia is difficult to predict and the quantification of metamorphopsia may be helpful to identify patients who may benefit the most from surgery.

Recently, the emergence of digital health solutions, particularly smartphone-based apps, has introduced new possibilities for accessible and practical tools to assess metamorphopsia. The Alleye app (Oculocare Medical Inc), a smartphone-based hyperacuity test (SHT) with Food and Drug Administration 510 (k) clearance, is the first mobile app to provide a quantitative metamorphopsia result. It is already in use for monitoring the progression of metamorphopsia in retinal diseases such as age-related macular degeneration [4,5]. Compared with the conventional printed M-chart for metamorphopsia quantification, the SHT provides a unified score that supports longitudinal monitoring, examines more retinal axes for a comprehensive assessment, and incorporates gamification to improve patient adherence [4,5]. This may make it a practical tool for continuous care, especially as patients with ERM require long-term monitoring with frequent checkups. The SHT's remote testing capability may reduce the need for in-person visits and support telemedicine pathways, aligning with modern trends in patient-centered health care. Its relevance in the presence of ERM has yet to be evaluated.

This study represents the first attempt to evaluate the SHT in patients with idiopathic epiretinal membranes (iERM) by correlating the SHT (Alleye app) to the conventional printed chart (M-chart) for metamorphopsia quantification before and after iERM surgery. It aims to explore whether the SHT can serve as a reliable, efficient tool for metamorphopsia measurement in a clinical setting. In addition, the study also investigated the correlation between the metamorphopsia scores and retinal OCT biomarkers as well as subjective perception of metamorphopsia.

Methods

Ethical Considerations

The study was approved by the ethics committee of the City of Vienna (approval no. EK21-027-0321) and the Austrian Agency for Health and Food Safety. All research activities were conducted in accordance with institutional and national guidelines and complied with the Declaration of Helsinki and the Good Clinical Practice guidelines of the European Union. This study was registered at ClinicalTrials.gov with the identification number NCT05138315. Written informed consent was obtained from all participants before their enrollment in the study. All participants were fully informed about the nature of the study, and they had the opportunity to opt out at any time. Participant data were pseudonymized and deidentified to ensure confidentiality. Strict protective measures were in place to safeguard all participant information throughout the study. No financial compensation was provided to participants for their involvement in the study and participation was entirely voluntary. No images of identifiable individuals are included in the manuscript.

Study Design and Patients

This prospective observational study was conducted at the department of ophthalmology of the Hanusch Hospital, Vienna, Austria. The study included 30 eyes of 30 patients who met the inclusion criteria and were scheduled for membrane peeling with vitrectomy for iERM. Eligibility criteria included the presence of iERM, age above 18 years, written informed consent, best distance-corrected visual acuity (DCVA) ≤1.0 logMAR and metamorphopsia detected on the Amsler grid. Patients diagnosed with other macular disorders including age-related macular degeneration or participants who had undergone previous intraocular surgery except for uncomplicated cataract surgery were excluded from the study.

Preoperative examinations and the 3 months postoperative follow-up were performed by the team of the Vienna Institute for Research in Ocular Surgery, including an ophthalmologist (CL). Study-related ophthalmic examinations included slit lamp biomicroscopy, retinal examination, DCVA Early Treatment Diabetic Retinopathy Study (ETDRS), metamorphopsia testing (Amsler grid, SHT, and M-charts), and spectral-domain (SD)-OCT with the CE-certified Spectralis OCT (Heidelberg Engineering). The SD-OCT images were obtained using the fast scan protocol (25 sections, 240 µm, 30° angle, 0.75 D focus, AUTO 71 sensitivity, 100% IR power, OCT Volume mode, and high-speed rate of 8.8/s). Apart from these examinations, slit lamp biomicroscopy and SD-OCT were routinely performed in the outpatient department 1 week and 1 month after surgery.

Surgical Procedure

All patients included into this study were operated by the same surgeon. Surgery included a 23-gauge pars plana vitrectomy and membrane peeling in all cases. For visualization of the ERM and internal limiting membrane (ILM), chromovitrectomy was performed with a trypan blue and brilliant blue G-based dye (MembraneBlue Dual; DORC). The ERM peeling was performed using an end-gripping forceps. Peeling of the ILM was performed as a second step in all cases where it had not been removed en bloc with the ERM. A restaining has been performed in all patients to identify and peel ILM residues. If air tamponade was needed, a complete fluid-to-air exchange was performed at the end of surgery and patients were advised to postoperative prone positioning for 24 hours. Gas tamponade (SF6 or C3F8) was used only in cases with coexistent peripheral retinal breaks.

Metamorphopsia Tests

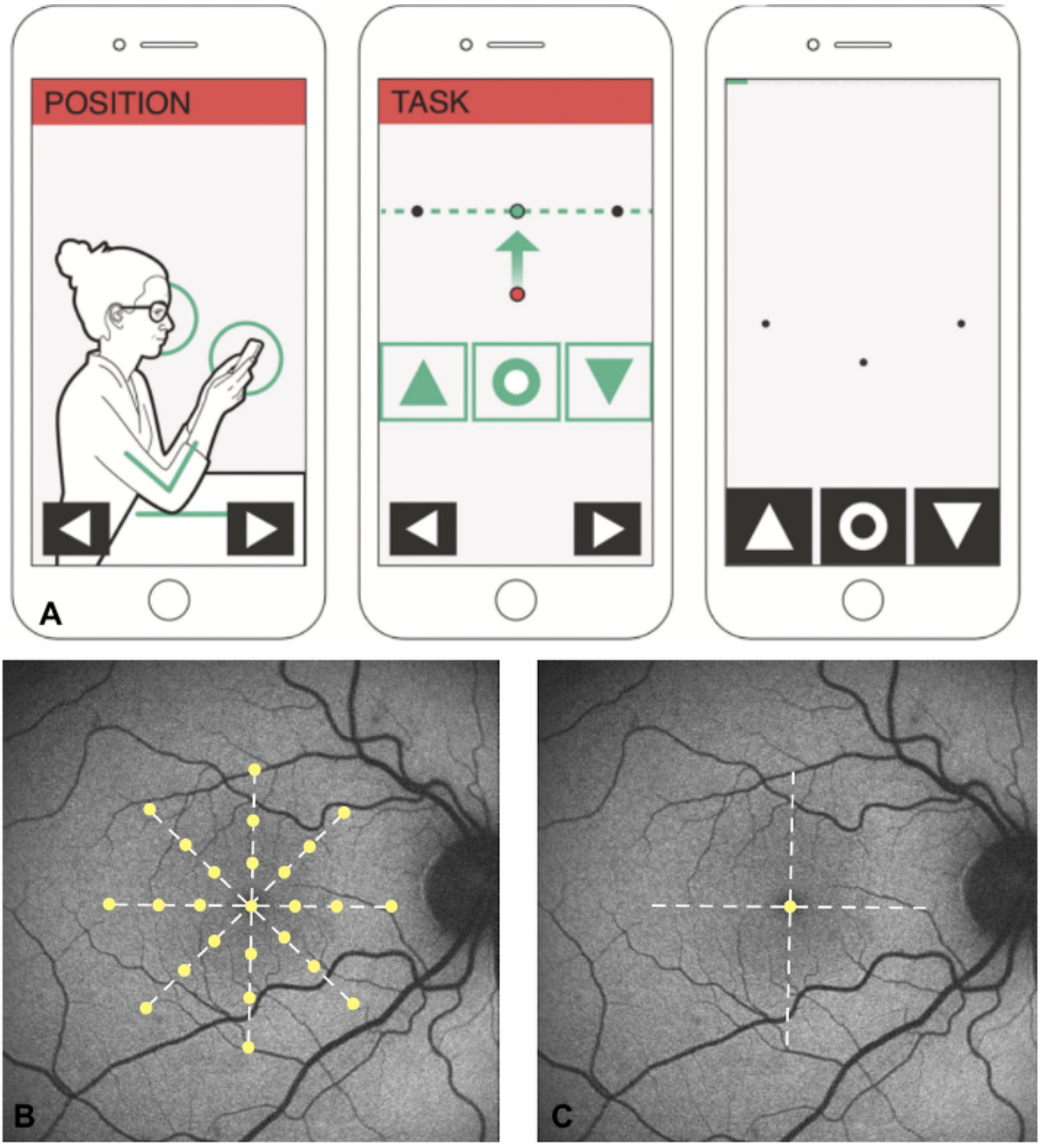
The main outcome variables in this study are the preoperative and 3 months postoperative mean SHT scores and mean M-chart scores. Both the M-chart and the SHT were consecutively performed 3 times before surgery and 3 months after surgery. The mean scores were used for statistical analysis and reliability of the 3 test repetitions was assessed. Detection of metamorphopsia on the Amsler grid was also included. The Amsler test was performed only once before and after surgery, as it provides a simple yes or no answer and was used as an inclusion criterion for the study. During the metamorphopsia examinations the eye not being tested was covered. In case patients needed reading glasses for near vision, they were asked to wear their own glasses or received a suitable near addition. All metamorphopsia examinations were tested with about 30-cm distance from the face and iPad rotation was standardized to a horizontal alignment.

For the Amsler examination, the participants were asked to fixate on the central dot of the grid and evaluate whether the lines are straight and parallel and whether the squares are regular and equal in size. If any of these characteristics were mentioned, the Amsler test was positive.

The M-chart was performed 3 times both in horizontal (MH score) and vertical (MV score) planes. The examiner alternated between the 2 positions to get more objective scores. The test has 19 dotted lines with dot intervals between 0.2º and 2.0º visual angles. The patients were shown these dotted lines beginning with 0º until the minimum visual angle needed to cause the metamorphopsia to disappear being the score.

The SHT (Alleye app) software can be downloaded on Apple iOS devices or Google Play. The SHT was consecutively performed 3 times and the testing procedure before and after surgery was the same. The patients received an oral explanation of the examination aided by the “training” option on the app. The eye not being tested was covered. The patient held the iOS device (iPad) and the task of the SHT was to align the central dot on an imaginary straight line between the paracentral points by tapping the arrow keys (Figure 1). This assignment was repeated 12 times in 4 different axes until the test was completed and the patient received a score between zero and 100. A higher score should be indicative of less metamorphopsia in contrast to the M-chart where a higher score should indicate more pronounced metamorphopsia. The SHT covers 4 axes with fixation dots at different positions while the M-chart covers 2 axes with a central single fixation dot (Figure 1).

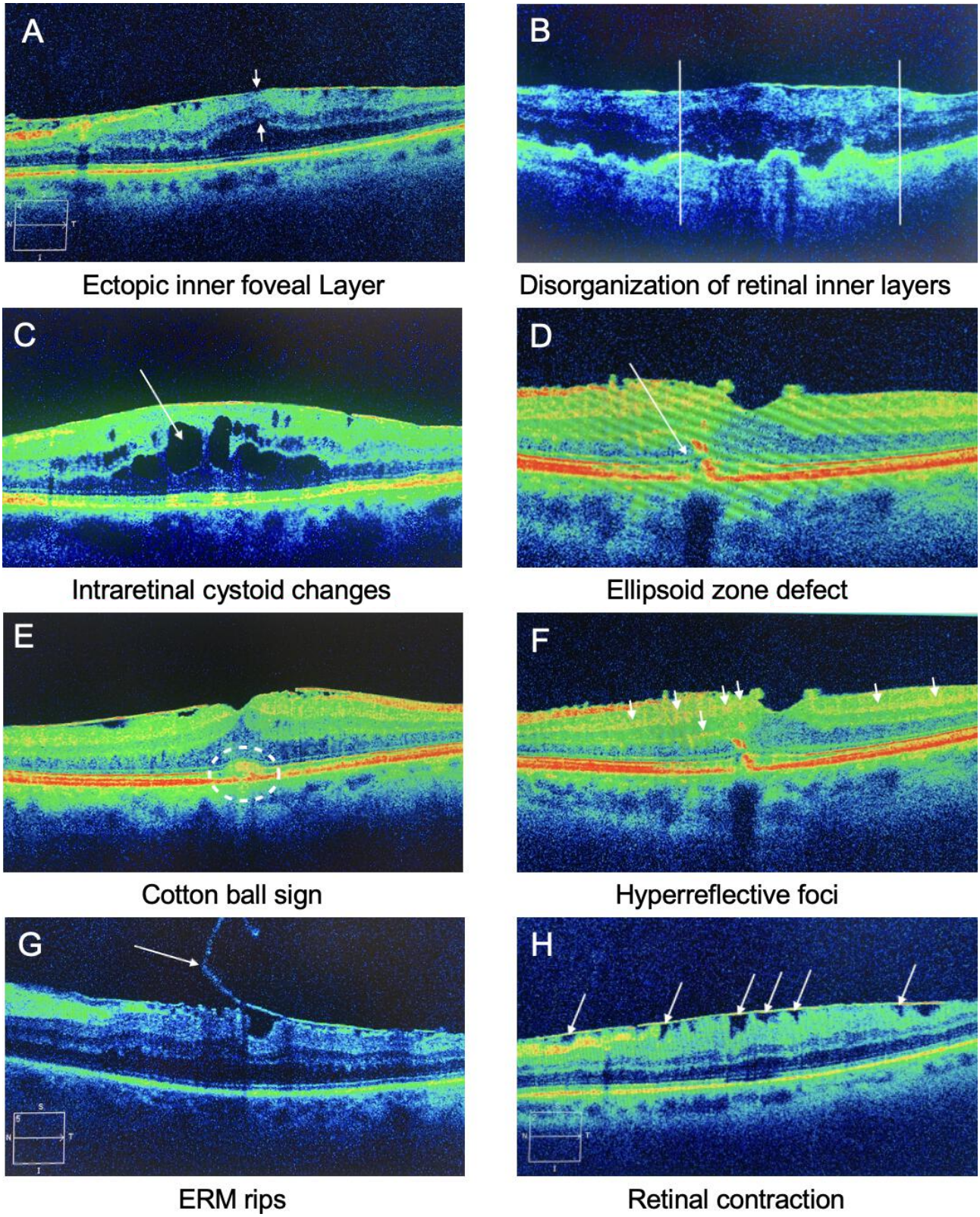
Figure 1. Correct positioning of the smartphone-based hyperacuity test (SHT) and alignment of the central dot on an imaginary straight line between the paracentral points (A) [6]. Four axes examined 3 times by the SHT with the target dot placed at different positions (B) compared with the M-chart with the horizontal and vertical axis and a central fixation dot (C) [7].



SD-OCT Biomarker Evaluations

As a secondary outcome this study evaluated the correlation of metamorphopsia scores and OCT biomarkers. OCT biomarkers are specific changes in retinal morphology with possible influence on postsurgical outcome. The preoperative SD-OCT images were evaluated by 2 independent graders (AD and SA) who were ophthalmology residents trained for OCT diagnosis. The readings of the 2 readers were compared and in case of discrepancies, the third reader, an experienced retina specialist made a final decision. The images were screened regarding the presence or absence of the following SD-OCT biomarkers: ectopic inner foveal layer, disorganization of retinal inner layers, intraretinal cystoid changes, ellipsoid zone defect, cotton ball sign, hyperreflective (HR) foci, ERM rips, and retinal contraction (Figure 2). Central macular thickness was also included in the analysis.

Figure 2. Spectral-domain optical coherence tomography (SD-OCT) biomarkers A–H (pictures taken and modified by Amon D). (A) Ectopic inner foveal layer is the presence of a continuous hypo- or hyperreflective lamina expanding from the inner nuclear layer and inner plexiform layer over the foveal zone as indicated by the arrows. (B) Disorganisation of retinal inner layers is present when the borders between the inner retinal layers are not recognisable as indictated between the two lines. (C) Intraretinal cystoid changes are fluid accumulations within the retina that can be recognised as hyporeflective spaces on OCT scans as indicated by the arrow. (D) The ellipsoid zone (EZ) is a hyperreflective region built between the interface of inner and outer photoreceptor segments and disruption can be seen as discontinuation of the hyperreflective EZ band as indicated by the arrow. (E) Cotton ball sign is defined as a round or diffuse hyperreflective area between the EZ and the interdigitation zone within the centre of the fovea as indicated by the circle. (F) Hyperreflective foci appear as small, highly reflective dots scattered within different layers of the retina as indicated by the arrows. (G) Epiretinal membrane rips are depicted as a torn edge of the ERM with a scrolled flap as indicated by the arrow. (H) Retinal contraction appears as wrinkling of the underlying retina caused by contraction of the ERM as indictaed by the arrows. ERM: epiretinal membrane.



Subjective Perception of Metamorphopsia

This study additionally tried to quantify the subjective perception of metamorphopsia using a telephone questionnaire. We postoperatively asked patients to grade their pre- and postoperative perception of metamorphopsia on a scale ranging from "0" indicating no distortion to "5" indicating severe distortion. Correlations between the subjective grades and standardized metamorphopsia scores were evaluated.

Statistical Analysis

Statistical analyses were conducted using SPSS (version 28; IBM SPSS Statistics). Level of significance was defined as a *P* value below .05. The Shapiro-Wilk test was used to test for normality. Reproducibility for the SHT and M-chart test was evaluated using intraclass correlation coefficients (ICCs) between 3 consecutive measurements pre- and postoperatively. We used the interpretation by Koo and Li [8] suggesting that ICC values below 0.5 are indicative of poor reliability, values between 0.5 and 0.75 are indicative of moderate reliability, values between 0.75 and 0.9 are indicative of good reliability, and values greater than 0.90 are indicative of excellent reliability. Correlations between the mean (mean of 3 examinations) SHT scores, mean MH, mean MV, the sum of mean MH and mean MV (MH+MV mean), and DCVA were calculated with a bivariate correlation. Paired 2-tailed *t* tests were applied to analyze the potential difference between preoperative and postoperative SHT and M-chart scores as well as DCVA. The *t* tests and bivariate correlation were used for the evaluation of the relationship between metamorphopsia scores and postoperative positive Amsler test, OCT biomarkers, and subjective metamorphopsia scores. Correlation coefficients were calculated using the Pearson correlation (*r*). Classification of the magnitude of correlation by Wuensch and Evans [9] was used interpreting an *r* value below 0.20 as very weak, *r* values between 0.20 and 0.39 as week, *r* values between 0.40 and 0.59 as moderate, *r* values between 0.60 and 0.79 as strong, and *r* value above 0.80 as very strong correlation.

Results

Demographic Data and Visual Acuity

A total of 30 patients were included into this study. Three of the 30 patients who had undergone preoperative examinations and surgery for iERM were lost to postoperative follow-up as they did not wish to participate in the postoperative study examinations and were therefore excluded from the final analysis. The mean age of our patient cohort was 71.2 (SD 8.2) years (Table 1). Regarding lens status, 12 patients were pseudophakic preoperatively and 15 were phakic preoperatively. Of the 15 preoperative patients with phakia, 12 patients underwent combined phacovitrectomy. Surgery resulted in a mean DCVA change of 15.3 (SD 9.3) ETDRS letters, with a minimum of −7 letters and a maximum change of 35 letters (*P*<.001; Table 2). The mean preoperative DCVA was 62.3 (SD 11.6) ETDRS letters compared with the mean postoperative DCVA with 77.6 (SD 8.2) letters. No significant correlations between the metamorphopsia scores and change in DCVA were found (Table 3).

Table . Demographic data (N=27).

Characteristics	Values
Age (years)	
Mean (SD)	71.2 (8.2)
Maximum	88
Minimum	54
Sex, n (%)	
Female	9 (33)
Male	18 (66)
Eye, n (%)	
Right eye	12 (44)
Left eye	15 (55)
Lens status, n (%)	
Preoperative phakia	15 (56)
Preoperative pseudophakia	12 (44)
Combined phacovitrectomy	12 (44)

Table . Visual acuity outcomes (Early Treatment Diabetic Retinopathy Study letters).

	Preoperative DCVA[a]	Postoperative DCVA	Change in DCVA
Mean (SD)	62.3 (11.6)	77.6 (8.2)	15.3 (9.3)
Maximum	82	90	35
Minimum	40	60	–7

[a]DCVA: distance-corrected visual acuity.

Table . Correlations between metamorphopsia scores and change in distance-corrected visual acuity.

	Pearson *r*	*P* value
SHT[a] mean preoperative	–0.17	.40
MH[b] mean preoperative	–0.16	.41
MV[c] mean preoperative	0.03	.87
MH+MV mean preoperative	–0.11	.59
SHT mean postoperative	–0.26	.18
MH mean postoperative	0.14	.49
MV mean postoperative	0.25	.22
MH+MV mean postoperative	0.20	.32

[a]SHT: smartphone-based hyperacuity test.

[b]MH: horizontal M-chart.

[c]MV: vertical M-chart.

Metamorphopsia Test Results

Change of Metamorphopsia After Surgery

Patients had significantly higher (r=0.69; P<.001) postoperative SHT scores compared with scores before surgery and a strong correlation between the 2 variables was found (Figure 3). The preoperative mean SHT score was 55.2 (SD 18.9) compared with 63.5 (SD 16.3) postoperatively resulting in a difference of 8.30 points. The improvement of M-chart scores, however, did not prove to be significant and the preoperative and postoperative values showed a weak correlation. The mean preoperative MH changed from 0.58 (SD 0.39) to 0.43 (SD 0.42) (r=0.37; P=.06) with a difference of 0.15, mean preoperative MV from 0.50 (SD 0.24) to 0.43 (SD 0.36) (r=0.18; P=.36) with a difference of 0.07, and mean preoperative MH+MV from 1.08 (SD 0.52) to 0.87 (SD 0.73) (r=0.25; P=.20) with a difference of 0.21 (Table 4).

Figure 3. Scatter plot depicting dots above the 45° line indicate patients who had improved smartphone-based hyperacuity test scores after epiretinal membrane surgery.

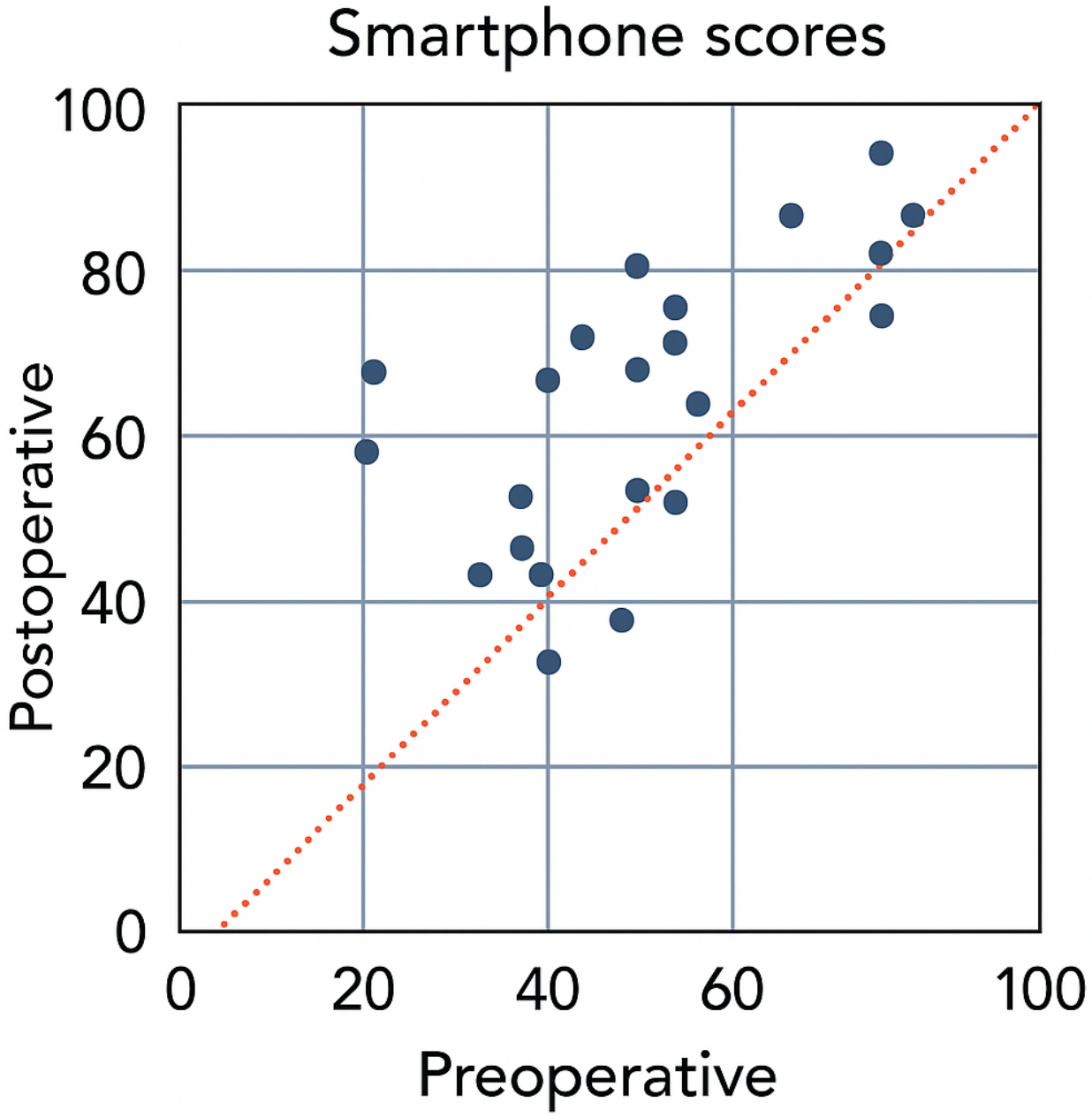


Table . Correlations between metamorphopsia scores.

Pairs of metamorphopsia scores	Mean (SD)	Pearson *r*	*P* value
Pair 1			
SHT[a] mean preoperative	55.22 (18.85)	0.69	<.001
SHT mean postoperative	63.52 (16.26)		
Pair 2			
MH[b] mean preoperative	0.58 (0.39)	0.37	.06
MH mean postoperative	0.43 (0.42)		
Pair 3			
MV[c] mean preoperative	0.50 (0.24)	0.18	.83
MV mean postoperative	0.43 (0.36)		
Pair 4			
MH+MV mean preoperative	1.08 (0.52)	0.25	.21
MH+MV mean postoperative	0.87 (0.73)		
Pair 5			
SHT mean preoperative	55.22 (18.85)	−0.18	.369
MH mean preoperative	0.58 (0.39)		
Pair 6			
SHT mean preoperative	55.22 (18.85)	0.01	.97
MV mean preoperative	0.50 (0.24)		
Pair 7			
SHT mean preoperative	55.22 (18.85)	−0.13	.51
MH+MV mean preoperative	1.08 (0.52)		
Pair 8			
MH mean preoperative	0.58 (0.39)	0.32	.10
MV mean preoperative	0.50 (0.24)		
Pair 9			
SHT mean postoperative	63.52 (16.26)	−0.14	.48
MH mean postoperative	0.43 (0.42)		
Pair 10			
SHT mean postoperative	63.52 (16.26)	−0.20	.32
MV mean postoperative	0.43 (0.36)		
Pair 11			
SHT mean postoperative	63.52 (16.26)	−0.18	.37
MH+MV mean postoperative	0.87 (0.73)		
Pair 12			
MH mean postoperative	0.43 (0.42)	0.74	<.001
MV mean postoperative	0.43 (0.36)		

[a]SHT: smartphone-based hyperacuity test.

[b]MH: horizontal M-chart.

[c]MV: vertical M-chart.

Reproducibility of Metamorphopsia Tests

The ICCs of all the metamorphopsia measurements showed good reliability (Table 5). The horizontal and vertical M-chart tests achieved higher scores than the SHT showing excellent reliability before and after surgery.

Table . Intraclass correlation coefficients of metamorphopsia tests.

Metamorphopsia test	Intraclass correlation coefficient
SHT[a] preoperative	0.87
MH[b] preoperative	0.96
MV[c] preoperative	0.94
SHT postoperative	0.82
MH postoperative	0.97
MV postoperative	0.93

[a]SHT: smartphone-based hyperacuity test.

[b]MH: horizontal M-chart.

[c]MV: vertical M-chart.

Correlations Between SHT and M-Chart

Correlations between the SHT and the M-chart were found to be very weak and not significant preoperatively as well as postoperatively. The correlation coefficients of the metamorphopsia scores and the corresponding levels of significance are listed in Table 4.

Correlation Between Metamorphopsia Scores and Amsler Grid

In 7 patients, metamorphopsia was not detected on the Amsler grid after surgery compared with 20 patients who still showed a positive postoperative Amsler test. The group of patients with a positive Amsler test after surgery showed significantly (r=−0.46; P=.02) worse postoperative SHT scores than the group without (Table 6). The mean SHT score of the patients with metamorphopsia detected on the Amsler grid after surgery was 59.2 (SD 15.3) compared with 75.7 (SD 12.9) of the patients with a negative Amsler test resulting in a difference of 16.7 between these groups. The group of patients with a postoperative positive Amsler test also showed significantly worse MV (r=0.39; P=.04) and MH+MV (r=0.39; P=.047) scores while the MH score was not significantly different between the groups (r=0.33; P=.09).

Table . Correlations between metamorphopsia scores and Amsler grid.

Postoperative	Amsler +	N	Mean (SD)	Pearson *r*	*P* value
SHT[a] score	Negative	7	75.9 (12.9)	−0.46	.02
	Positive	20	59.2 (15.3)		
MH[b] score	Negative	7	0.2 (0.3)	0.33	.09
	Positive	20	0.5 (0.4)		
MV[c] score	Negative	7	0.2 (0.3)	0.39	.04
	Positive	20	0.5 (0.4)		
MH+MV score	Negative	7	0.4 (0.5)	0.39	.047
	Positive	20	1.0 (0.7)		

[a]SHT: smartphone-based hyperacuity test.

[b]MH: horizontal M-chart.

[c]MV: vertical M-chart.

Correlation Between Metamorphopsia Scores and SD-OCT Biomarkers

Central macular thickness (r=−0.44; P=.02; r=−0.46; P=.02) and intraretinal cysts (r=−0.72; P<.001; r=−0.65; P<.001) proved to be significantly associated with pre- and postoperative SHT scores and HR foci showed a significant correlation with the postoperative SHT score (r=0.44; P=.02). Regarding the M-chart scores, the disorganization of retinal inner layers (r=0.68, P<.001; r=0.58, P=.002) and ellipsoid zone defect (r=0.49, P=.01; r=0.48, P=.01) showed significant associations with the preoperative vertical M-chart score (r=0.68; P<.001) and the preoperative MH+MV score (r=0.58; P=.002). The other biomarkers were not significantly associated with any of the metamorphopsia scores and correlations proved to be very weak to weak. Detailed results of the biomarker readings can be found in the Multimedia Appendix 1.

Correlation Between Metamorphopsia Scores and Subjective Metamorphopsia Perception

Out of 21 patients surveyed, 19 reported a subjective improvement in metamorphopsia after surgery. The mean subjective scores improved by 2.2 points from 3.7 (SD 0.9) preoperatively to 1.5 (SD 1.3) postoperatively, showing a significant correlation (r=0.55; P=.01;). The preoperative (r=0.58; P=.01) and postoperative (r=0.81; P<.001) subjective metamorphopsia scores correlated significantly with postoperative M-Chart scores, but not significantly with SHT scores. Patients with postoperative positive Amsler tests reported significantly more severe subjective metamorphopsia scores (r=0.49; P=.02).

Discussion

Principal Findings

The aim of this study was to estimate the correlation between the SHT (Alleye app) and a conventional printed chart (M-chart) for metamorphopsia quantification before and after vitrectomy with membrane peeling for patients with iERM.

The SHT has mainly been studied for other retinal diseases and this study is the first to ever investigate the SHT in patients with iERM. It is the first mobile app to provide a quantitative metamorphopsia result. Contrary to the vertical and horizontal scores of the M-chart, the SHT provides a single quantitative score including 4 axes instead of 2.

The SHT scores significantly (r=0.69; P<.001) improved from 55.2 (SD 18.9) preoperatively to 63.5 (SD 16.3) after surgery resulting in an improvement of 8.3 points. The mean postoperative M-chart scores also improved after surgery, but were not statistically significant. The significant improvement of postoperative SHT scores suggests that patients with metamorphopsia due to iERM may benefit from surgery and the SHT can potentially be a software to quantify these results. Clinicians routinely do not test for metamorphopsia, despite its relevance for surgical decisions. Kinoshita et al [3,10] found a significant correlation between preoperative and postoperative M-chart scores proposing the consideration of surgery before severe worsening of metamorphopsia and a stage where it cannot be completely resolved. Testing macular function with visual acuity alone is often insufficient, as it fails to capture metamorphopsia, a subjective perception that is challenging to measure objectively. In our study, both the SHT and the M-chart demonstrated good reliability, as shown by their ICCs. However, the SHT may offer a more comprehensive assessment by covering 4 retinal axes, whereas the M-chart evaluates only 2. It is important to note that this study did not aim to test the superiority of the SHT over the M-chart but was a first attempt to evaluate the SHT in patients with iERM aiming to identify a practical tool that enables efficient and reliable measurement of metamorphopsia in clinical practice. The tests provide complementary but distinct insights into the severity of metamorphopsia, and their scores should not be considered directly comparable. The SHT is primarily used for detecting macular edema. This was the first attempt to evaluate the SHT in patients with ERM, providing initial insights into its potential use for this condition. In addition, the SHT, being a more practical and digital tool, could offer significant advantages in clinical practice. The SHT's ability to provide a unified score, enable remote monitoring, and incorporate gamification may improve patient engagement and reduce unnecessary clinic visits.

Of interest, the results of our analysis showed a very weak and not significant association between the SHT and the M-chart. Since the M-chart includes only 2 axes, horizontal and vertical, while the SHT includes the horizontal, vertical, and oblique axes, there is a discrepancy between the examined retinal axes. A key limitation of this study is the inability to analyze individual horizontal, vertical, and oblique axes measured by the SHT, which could allow for a more direct comparison with the M-chart, which evaluates horizontal and vertical axes separately. To minimize the discrepancy between the examined axes, we combined the 2 M-chart axes and calculated the sum of the mean vertical M-chart score and mean horizontal M-chart score. The 2 axes included in 1 score, however, also did not significantly correlate with the SHT score. Analyses of the SHT's individual axes might help clarify whether the weak correlation observed between the 2 tests is due to differences in the retinal areas examined. However, this was not feasible due to the proprietary design of the SHT, which provides only a combined score rather than axis-specific data. This limitation has been communicated to the developers, and we recommend incorporating functionality for individual axis analysis in future updates to enhance the SHT's scientific use. Despite this constraint, studies have demonstrated the reliability and practicality of the SHT. Studies reported high ICCs, diagnostic reliability for monitoring macular function, good patient adherence and usability for remote monitoring, and a low threshold for use, making it accessible and effective in real-world settings [4,11-13]. These advantages position the SHT as a practical alternative for assessing metamorphopsia in clinical practice, even without detailed axis-specific data. In addition, the SHT's combined score approach may simplify the interpretation of metamorphopsia as a general symptom of macular dysfunction, supporting its use in patient-centered care. Nevertheless, the mapping of distorted vision to a specific axis may additionally be of use since the horizontal M-chart score tends to improve to a larger extent than the vertical score that arises in later stages suggesting that once the vertical distortion is present, it is less likely to resolve compared with horizontal metamorphopsia [10]. Our study as well as the study by Kinoshita et al [3,10] showed that the baseline MH score was higher than the MV score. The vertical plasticity may be greater than that for the horizontal retina since the axons of the retinal ganglion cells run horizontally rather than vertically in the posterior pole in addition to the optic disc that might also reduce horizontal displacement in the posterior pole. Since vertical contraction is perceived as horizontal metamorphopsia and vice versa, the horizontal M-chart scores might be higher than the vertical scores [14,15]. The prognostic properties of individual metamorphopsia axes may be of importance in advising patients and giving them a realistic prognosis for postoperative outcomes. Schmid et al [4] claimed that when using the SHT, patients actively need to align a central point on an imaginary straight line and the outer points remain stable in the paracentral visual

field thereby ensuring that patients fixate on the moving dot and avoiding saccades to the outer areas. This should ensure a proper fixation for metamorphopsia testing. In passive tests such as the M-chart or the Amsler grid, where patients are shown different lines or grids and asked about their perception of them, fixation on paracentral areas is more likely, which may lead to the conclusion of lower reproducibility. In our study, however, all metamorphopsia tests showed a good reliability (Table 5) and the M-chart even achieved higher scores than the SHT proving excellent reliability before and after surgery. This underlines the findings by Matsumoto et al [16] reporting good reliability of M-charts and intraindividual variation to be within 1 line (±0.1 score).

Regarding associations of metamorphopsia scores with the Amsler grid, almost a third of the patients had improved metamorphopsia after surgery when testing with the Amsler grid. The group of patients with a positive Amsler test after surgery showed significantly worse postoperative SHT scores (r=−0.46; P=.02), MV scores (r=0.40; P=.04), and MH+MV scores (r=0.39; P=.047) than the group without. The results suggest that a persistent positive Amsler test may be a reliable predictor of poor SHT scores and therefore outcome.

Postoperative mean visual acuity improved by 3 lines and no significant correlation between DCVA and metamorphopsia was found which is in accordance with literature [3,17-19]. Lens status revealed that 12 patients were pseudophakic preoperatively, while 15 were phakic, of whom 12 underwent combined phacovitrectomy. Since visual acuity was not the main focus of this study, the lens status was not further emphasized in the analysis. Metamorphopsia seems to be independent of visual acuity and an important symptom contributing to quality of vision.

Regarding the OCT biomarkers, a larger central macular thickness, the presence of intraretinal cysts, and HR foci were significantly associated with postoperative SHT scores. These findings suggest that these biomarkers may have prognostic properties for metamorphopsia outcomes after ERM surgery. This is particularly relevant in clinical decision-making, as identifying OCT biomarkers associated with postoperative visual distortion could help guide surgical timing and set realistic patient expectations. To validate our findings, larger studies and longer follow-up periods are necessary. These studies should further investigate the relationship between objective metamorphopsia scores, OCT biomarkers, and subjective metamorphopsia perception. Such research could refine the prognostic use of OCT findings and contribute to the development of comprehensive assessment tools that integrate objective measures with subjective experiences. Ultimately, this could lead to more tailored surgical interventions and improved patient outcomes in ERM management.

During the postoperative follow-up visits of our study, we frequently examined content patients stating that their visual distortion had highly improved and asking whether their scores had improved in an objective manner. This observation underlined the importance of an individual's perception of vision. Our study demonstrated that subjective perception of metamorphopsia in patients with iERM can improve with surgery and patients with a higher preoperative degree of metamorphopsia also suffer from more severe metamorphopsia postoperatively. Detailed results on the subjective outcomes are given in Multimedia Appendices 2 and 3.

An important aspect to consider with the SHT is the possibility for patients to quantitatively monitor the progression of metamorphopsia at home in a comfortable setting and compare their results with their last examinations. Interactive tests and the gamification of home-monitoring tasks can lead to a higher patient motivation to take ownership of their eye health as well as better protection of sight [20,21]. Smartphone apps can also serve as a tool for patients to remotely view their health record information [22]. The newest version of the SHT is designed to serve as a digital companion for patients to have a better overview of their retinal disease and keep them motivated to follow the treatment program. The novel features of the SHT enable patients and clinicians to collect data, manage appointments, and document diagnoses, medical results, visual acuity, injections, and subjective visual impairment [23]. Since many people already own a smartphone, the implementation of the SHT is rather straightforward. In our study, we included a large age range between 54 and 88 years and none of the patients needed extra help to perform the SHT after a short introduction at the beginning of the test. It is important, however, that patients who do not qualify for home monitoring are not disadvantaged and do not receive insufficient care for their eye health. Research suggests that smartphone-based vision monitoring is accessible across diverse population groups with varying levels of digital proficiency and social deprivation, making it a viable and reliable tool for monitoring clinical progression [24].

Currently, no established thresholds exist to define clinically significant changes in SHT scores or to determine when patients should seek medical advice. Further studies are necessary to address these gaps and to establish evidence-based benchmarks for the SHT's use in clinical decision-making. Previous research has provided M-chart–based thresholds for surgical indications; however, these cutoff values have not been adopted in clinical practice. Kinoshita et al [3] suggested that preoperative MH scores between 0.5 and 1.7 or MV scores between 0.5 and 0.9 could indicate the need for surgery. While strict cutoff values may be challenging in individual cases, quantifying metamorphopsia can enhance the decision-making processes. The SHT can complement traditional tools by providing additional, objective data to support surgical planning, facilitate triaging, and monitor disease progression.

Compared with the conventional printed M-chart for metamorphopsia quantification, the SHT offers several advantages that align with modern, patient-centered health care approaches. The SHT provides a unified score that supports longitudinal monitoring and examines more retinal axes, offering a more comprehensive assessment of metamorphopsia. In addition, the incorporation of gamification enhances patient engagement and adherence, transforming a clinical task into an interactive and motivating experience for patients. These features may make the SHT a practical tool for continuous care, particularly for patients with ERM who require long-term monitoring with frequent checkups. Patients with ERM undergo follow-ups spanning several years, with frequent visits to

monitor disease progression or assess the need for surgical intervention. A reliable and easily accessible tool such as the SHT could serve as a valuable monitoring score to detect changes or dynamics in metamorphopsia over time. Its capability for remote testing can further reduce the need for in-person visits, supporting telemedicine pathways and addressing the increasing demand for accessible health care solutions. Moreover, the SHT can be particularly beneficial for patients in underserved areas, those with mobility limitations, or in regions with limited physician availability, ensuring broader access to effective disease monitoring. While the SHT has demonstrated its use in other retinal conditions, its relevance in the context of ERM has yet to be fully evaluated. These attributes suggest that the SHT could play a key role in the evolution of digital health tools for monitoring metamorphopsia and guiding complex surgical decision-making in patients with ERM.

This study represents a foundational step in evaluating the SHT's app in ERM, emphasizing the need for further research to validate its clinical relevance and establish its role in patient-centered ophthalmologic care.

Limitations

Limitations of this study include the small sample size and the limited follow-up time as metamorphopsia tends to improve for a longer period of about 12 months after ERM surgery and a small number of patients is a limitation for the validity of the test [3,10]. To date, it has unfortunately not been possible to retrieve the results of the individual axes of the SHT separately as it would be of great interest to depict what the diagonal planes measure and calculate their correlations with the horizontal and vertical planes of the M-Chart. Without this breakdown of SHT results, the hypothesis that the discrepancy in examined retinal axes may be a reason why the scores of the 2 tests did not significantly correlate with each other cannot be fully proven. Due to the exploratory nature of this study, we could not perform a formal sample size calculation. Another confounder of this study may be patient motivation and character as well as a learning effect with repeated testing. The potential influence of learning effects on repeated testing with the SHT and M-chart is an important consideration for interpreting the study findings. While we assessed test-retest reliability for both tools, we did not explicitly analyze learning effects. Previous research on the SHT, including a study by Faes et al [12], demonstrated its excellent usability with a median system usability score of 90. This indicates that most users found the app intuitive and easy to learn, suggesting that the influence of learning on SHT results may be minimal. Furthermore, the digital nature of the SHT, combined with its interactive and standardized testing process, reduces the likelihood of variability due to user fatigue or concentration, issues that are more likely to affect manual tools such as the M-chart. Incorporating design elements that prioritize user-friendly interfaces, as seen with the SHT, can help maintain adherence and reduce the impact of learning effects in mobile health tools. Future studies could further investigate the influence of learning on SHT performance. As not all patients were pseudophakic after surgery, the improvement of DCVA may have been influenced by concomitant cataract surgery; however, visual acuity was not a main outcome of this study. Our study was conducted at a single center and all surgeries were performed by the same retina surgeon thereby increasing reproducibility.

Conclusions

The study was the first to ever investigate the SHT in patients with iERM. We showed that quantitative data provided by the SHT significantly improved after membrane peeling suggesting that patients with metamorphopsia due to iERM can benefit from surgery, and this application may potentially be a software to quantify metamorphopsia in patients with iERM. The metamorphopsia scores of the SHT showed a very weak and insignificant association with the M-chart scores. It would be of great interest to depict what the diagonal planes of the SHT measure and calculate their correlations with the horizontal and vertical planes of the M-chart to validate whether the poor correlation between the 2 tests may be explained by their discrepancy in examined retinal areas. The SHT may serve as a practical, accessible, and patient-centered tool for monitoring metamorphopsia, supporting long-term care and decision-making for patients with ERM, particularly in telemedicine and underserved settings. While this study lays the groundwork for future research, further studies including the breakdown of metamorphopsia axes of the SHT as well as a larger sample size and longer follow-up period are required to validate our results.

Authors' Contributions

All authors collectively contributed to the conception of the study. JH was responsible for study design, methodology, data interpretation, and review of the manuscript. CL operated all patients and was responsible for data recruitment. OF co-designed the study and contributed to the manuscript. DA was involved in literature search, data recruitment, data analysis, and drafted the manuscript. MR, CP, NB, and AS were responsible for patient and data recruitment. All authors read, contributed to, and approved the final manuscript. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Spectral-domain optical coherence tomography biomarker readings of the pilot study.

[DOCX File, 15 KB - periop_v8i1e60959_app1.docx]

Multimedia Appendix 2
Correlations between preoperative spectral-domain optical coherence tomography biomarkers and preoperative metamorphopsia scores.
[DOCX File, 16 KB - periop_v8i1e60959_app2.docx]

Multimedia Appendix 3
Correlations between preoperative spectral-domain optical coherence tomography biomarkers and postoperative metamorphopsia scores.
[DOCX File, 16 KB - periop_v8i1e60959_app3.docx]

References

1. Mao J, Wu H, Liu C, et al. Changes in metamorphopsia, visual acuity, and central macular thickness after epiretinal membrane surgery in four preoperative stages classified with OCT B-scan images. J Ophthalmol 2019;2019:7931654. [doi: 10.1155/2019/7931654] [Medline: 31316825]
2. Díaz-Valverde A, Wu L. To peel or not to peel the internal limiting membrane in idiopathic epiretinal membranes. Retina (Philadelphia, Pa) 2018 Sep;38 Suppl 1:S5-S11. [doi: 10.1097/IAE.0000000000001906] [Medline: 29068917]
3. Kinoshita T, Imaizumi H, Okushiba U, Miyamoto H, Ogino T, Mitamura Y. Time course of changes in metamorphopsia, visual acuity, and OCT parameters after successful epiretinal membrane surgery. Invest Ophthalmol Vis Sci 2012 Jun 14;53(7):3592-3597. [doi: 10.1167/iovs.12-9493] [Medline: 22589432]
4. Schmid MK, Thiel MA, Lienhard K, Schlingemann RO, Faes L, Bachmann LM. Reliability and diagnostic performance of a novel mobile app for hyperacuity self-monitoring in patients with age-related macular degeneration. Eye (Lond) 2019 Oct;33(10):1584-1589. [doi: 10.1038/s41433-019-0455-6]
5. Brucker J, Bhatia V, Sahel JA, Girmens JF, Mohand-Saïd S. Odysight: a mobile medical application designed for remote monitoring-a prospective study comparison with standard clinical eye tests. Ophthalmol Ther 2019 Sep;8(3):461-476. [doi: 10.1007/s40123-019-0203-9] [Medline: 31346977]
6. Faes L, Islam M, Bachmann LM, Lienhard KR, Schmid MK, Sim DA. False alarms and the positive predictive value of smartphone-based hyperacuity home monitoring for the progression of macular disease: a prospective cohort study. Eye (Lond) 2021 Nov;35(11):3035-3040. [doi: 10.1038/s41433-020-01356-2] [Medline: 33414531]
7. Portugal T. Figure 3: normal fundus autofluorescence: uniform grayish signal in the fundus and a marked dark appearance in the optic nerve (absence of RPE) and retinal vessels (absorption of fluorescence by hemoglobin and other blood contents). URL: https://amdbook.org/content/figure-3-normal-fundus-autofluorescence-uniform-grayish-signal-fundus-and-marked-dark-appear [accessed 2023-06-04]
8. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016 Jun;15(2):155-163. [doi: 10.1016/j.jcm.2016.02.012] [Medline: 27330520]
9. Wuensch KL, Evans JD. Straightforward statistics for the behavioral sciences. J Am Stat Assoc 1996 Dec;91(436):1750. [doi: 10.2307/2291607]
10. Kinoshita T, Imaizumi H, Miyamoto H, Katome T, Semba K, Mitamura Y. Two-year results of metamorphopsia, visual acuity, and optical coherence tomographic parameters after epiretinal membrane surgery. Graefes Arch Clin Exp Ophthalmol 2016 Jun;254(6):1041-1049. [doi: 10.1007/s00417-015-3147-3] [Medline: 26319984]
11. Schmid MK, Faes L, Bachmann LM, Thiel MA. Accuracy of a self-monitoring test for identification and monitoring of age-related macular degeneration: a diagnostic case-control study. Open Ophthalmol J 2018;12(19–28):19-28. [doi: 10.2174/1874364101812010019] [Medline: 29619125]
12. Faes L, Golla K, Islam M, et al. System usability, user satisfaction and long-term adherence to mobile hyperacuity home monitoring-prospective follow-up study. EYE (Lond) 2023 Mar;37(4):650-654. [doi: 10.1038/s41433-022-01959-x] [Medline: 35292773]
13. Gross N, Bachmann LM, Islam M, et al. Visual outcomes and treatment adherence of patients with macular pathology using a mobile hyperacuity home-monitoring app: a matched-pair analysis. BMJ Open 2021 Dec 23;11(12):e056940. [doi: 10.1136/bmjopen-2021-056940] [Medline: 34949632]
14. Arimura E, Matsumoto C, Okuyama S, Takada S, Hashimoto S, Shimomura Y. Retinal contraction and metamorphopsia scores in eyes with idiopathic epiretinal membrane. Invest Ophthalmol Vis Sci 2005 Aug;46(8):2961-2966. [doi: 10.1167/iovs.04-1104] [Medline: 16043872]
15. Kuroyanagi K, Gunji H, Kato H, Gekka T, Arai K, Ito Y. Use of M-charts® in the evaluation of metamorphopsia after surgical removal of epiretinal membrane. Jpn J Clin Ophthalmol 2010;64:1551-1554.
16. Matsumoto C, Arimura E, Okuyama S, Takada S, Hashimoto S, Shimomura Y. Quantification of metamorphopsia in patients with epiretinal membranes. Invest Ophthalmol Vis Sci 2003 Sep;44(9):4012-4016. [doi: 10.1167/iovs.03-0117] [Medline: 12939323]

17. Leisser C, Amon DL, Huemer JC, Findl O. Diagnostic reliability of optical coherence tomography biomarkers for postsurgical success in visual acuity in patients with idiopathic epiretinal membranes. Klin Monbl Augenheilkd 2023 Oct;240(10):1207-1213. [doi: 10.1055/a-1756-5243] [Medline: 35426108]
18. Crincoli E, Savastano MC, Savastano A, et al. New artificial intelligence analysis for prediction of long-term visual improvement after epiretinal membrane surgery. Retina (Philadelphia, Pa) 2023 Feb 1;43(2):173-181. [doi: 10.1097/IAE.0000000000003646] [Medline: 36228144]
19. Nowomiejska K, Oleszczuk A, Brzozowska A, et al. M-charts as a tool for quantifying metamorphopsia in age-related macular degeneration treated with the bevacizumab injections. BMC Ophthalmol 2013 Apr 15;13(1):13. [doi: 10.1186/1471-2415-13-13] [Medline: 23587218]
20. Razavi H, Baglin E, Sharangan P, et al. Gaming to improve vision: 21st century self‐monitoring for patients with age‐related macular degeneration. Clin Exp Ophthalmol 2018 Jul;46(5):480-484. [doi: 10.1111/ceo.13097] [Medline: 29131493]
21. Kaiser PK, Wang YZ, He YG, Weisberger A, Wolf S, Smith CH. Feasibility of a novel remote daily monitoring system for age-related macular degeneration using mobile handheld devices: results of a pilot study. Retina (Philadelphia, Pa) 2013 Oct;33(9):1863-1870. [doi: 10.1097/IAE.0b013e3182899258] [Medline: 23609122]
22. NHS digital. Online access to GP health records. 2025 Mar 5. URL: https://digital.nhs.uk/services/nhs-app/nhs-app-guidance-for-gp-practices/guidance-on-nhs-app-features/online-access-to-gp-health-records [accessed 2023-06-12]
23. Alleye. Alleye becomes a digital companion for patients with retinal disease. : Medium; 2023. URL: https://medium.com/@technology_11408/alleye-becomes-a-digital-companion-for-patients-with-a-retinal-disease-f98e234929c5 [accessed 2023-06-12]
24. Mendall J, Islam M, Wong K, et al. Digital exclusion, social deprivation, and clinical outcomes of patients undergoing hyperacuity home monitoring. Ophthalmol Ther 2024 Oct;13(10):2759-2769. [doi: 10.1007/s40123-024-01020-y] [Medline: 39181973]

Abbreviations

DCVA: distance-corrected visual acuity
ERM: epiretinal membrane
HR: hyperreflective
ICC: intraclass correlation coefficient
iERM: idiopathic epiretinal membrane
ILM: internal limiting membrane
MH: horizontal M-chart
MV: vertical M-chart
OCT: optical coherence tomography
SD-OCT: spectral-domain optical coherence tomography
SHT: smartphone-based hyperacuity test

Edited by N Rohatgi; submitted 27.05.24; peer-reviewed by G Piccoli, G Lim, G Kahraman, JM Teixeira; revised version received 04.12.24; accepted 04.03.25; published 17.04.25.

Please cite as:
Amon D, Leisser C, Schlatter A, Ruiss M, Pilwachs C, Bayer N, Huemer J, Findl O
Quantification of Metamorphopsia Using a Smartphone-Based Hyperacuity Test in Patients With Idiopathic Epiretinal Membranes: Prospective Observational Study
JMIR Perioper Med 2025;8:e60959
URL: https://periop.jmir.org/2025/1/e60959
doi:10.2196/60959

Original Paper

Enhancing Quadruple Health Outcomes After Thoracic Surgery: Feasibility Pilot Randomized Controlled Trial Using Digital Home Monitoring

Mahesh Nagappa[1*], MD; Yamini Subramani[1*], MD; Homer Yang[1*], MD; Natasha Wood[2*], MSc; Jill Querney[1*], MD; Lee-Anne Fochesato[1*], MSc; Derek Nguyen[1*], BMSc; Nida Fatima[1*], MBBS; Janet Martin[1*], MSc; Ava John-Baptiste[1*], MSc; Rahul Nayak[3*], MD; Mehdi Qiabi[3*], MD; Richard Inculet[3*], MD; Dalilah Fortin[3*], MD; Richard Malthaner[3*], MD

[1]Department of Anesthesia and Perioperative Medicine, London Health Sciences Centre and St. Joseph Health Care, Lawson Health Research Institute, Schulich School of Medicine and Dentistry, Western University, London, ON, Canada

[2]Department of Nursing, Thoracic Surgery, London Health Sciences Centre, Lawson Health Research Institute, Schulich School of Medicine and Dentistry, Western University, London, ON, Canada

[3]Department of Surgery, Division of Thoracic Surgery, London Health Sciences Centre and St. Joseph Health Care, Lawson Health Research Institute, Schulich School of Medicine and Dentistry, Western University, London, ON, Canada

[*]all authors contributed equally

Corresponding Author:
Mahesh Nagappa, MD
Department of Anesthesia and Perioperative Medicine
London Health Sciences Centre and St. Joseph Health Care, Lawson Health Research Institute, Schulich School of Medicine and Dentistry
Western University
339 Windermere Road
London, ON, N6A 5A5
Canada
Phone: 1 5196858500 ext 34436
Email: Mahesh.Nagappa@lhsc.on.ca

Abstract

Background: Surgical recovery after hospital discharge often presents challenges for patients and caregivers. Postoperative complications and poorly managed pain at home can lead to unexpected visits to the emergency department (ED) and readmission to the hospital. Digital home monitoring (DHM) may improve postoperative care compared to standard methods.

Objective: We conducted a feasibility study for a randomized controlled trial (RCT) to assess DHM's effectiveness following thoracic surgical procedures compared to standard care.

Methods: We conducted a 2-arm parallel-group pilot RCT at a single tertiary care center. Adult patients undergoing thoracic surgical procedures were randomized 1:1 into 2 groups: the DHM group and the standard of care (control group). We adhered to the intention-to-treat analysis principle. The primary outcome was predetermined RCT feasibility criteria. The trial would be feasible if more than 75% of trial recruitment, protocol adherence, and data collection were achieved. Secondary outcomes included 30-day ED visit rates, 30-day readmission rates, postoperative complications, length of stay, postdischarge 30-day opioid consumption, 30-day quality of recovery, patient-program satisfaction, caregiver satisfaction, health care provider satisfaction, and cost per case.

Results: All RCT feasibility criteria were met. The trial recruitment rate was 87.9% (95% CI 79.4%-93.8%). Protocol adherence and outcome data collection rates were 96.3% (95% CI 89.4%-99.2%) and 98.7% (95% CI 92.9%-99.9%), respectively. In total, 80 patients were randomized, with 40 (50%) in the DHM group and 40 (50%) in the control group. Baseline patient and clinical characteristics were comparable between the 2 groups. The DHM group had fewer unplanned ED visits (2.7% vs 20.5%; *P*=.02), fewer unplanned admission rates (0% vs 7.6%; *P*=.24), lower rates of postoperative complications (20% vs 47.5%, *P*=.01) shorter hospital stays (4.0 vs 6.9 days; *P*=.05), but more opioid consumption (111.6, SD 110.9) vs 74.3, SD 71.9 mg morphine equivalents; *P*=.08) compared to the control group. DHM also resulted in shorter ED visit times (130, SD 0 vs 1048, SD 1093 minutes; *P*=.48) and lower cost per case (CAD $12,145 [US $ 8436.34], SD CAD $8779 [US $ 6098.20] vs CAD $17,247 [US $11,980.37], SD

CAD $15,313 [US $10,636.95]; *P*=.07). The quality of recovery scores was clinically significantly better than the controls (185.4, SD 2.6 vs 178.3, SD 3.3; *P*<.001). All 37 patients who completed the intervention answered the program satisfaction survey questionnaires (100%; 95% CI 90.5%-100%). Only 36 out of 80 caregivers responded to the caregiver satisfaction questionnaires at the end of the fourth week post hospital discharge (47.7%; 95% CI 35.7%-59.1%). Health care providers reported a 100% satisfaction rate.

Conclusions: This pilot RCT demonstrates the feasibility of conducting a full-scale trial to assess DHM's efficacy in improving postoperative care following thoracic surgery. DHM shows promise for enhancing continuity of care and warrants further investigation.

Trial Registration: ClinicalTrials.gov NCT04340960; https://clinicaltrials.gov/study/NCT04340960

(JMIR Perioper Med 2025;8:e58998) doi:10.2196/58998

KEYWORDS

remote monitor; digital home monitoring; continuity of care; quadruple health outcomes; patient satisfaction; caregivers satisfaction; healthcare provider satisfaction; feasibility; RCT; thoracic surgery; postoperative monitoring; surgical recovery; perioperative medicine; patient care; questionnaire

Introduction

Recovery following surgical discharge poses significant challenges for patients and their caregivers. This challenge is compounded by the growing practice of discharging patients earlier after surgical procedures, intensifying the postoperative care demands. Moreover, the health care system often operates within a framework of fragmented and poorly integrated services, exacerbating the difficulties faced by patients transitioning from hospital to home after surgery, which can lead to complications and inadequately managed pain, resulting in returns to the hospital or visits to the emergency department (ED) [1-5].

Numerous studies underscore the critical role of postdischarge continuity of care in reducing ED visits and readmission rates (RRs) [6-11]. For instance, Shargall et al [12] successfully implemented an “Integrated Comprehensive Care” program involving allied health care professionals, significantly reducing 30-day RRs among thoracic surgery patients. Similar reductions in RRs have been attributed to patient education, well-coordinated discharge planning, physician follow-up, and in-home visits [13]. Data from Canada highlight that within the first 7 days following surgical discharge, 28.3% of ED diagnoses fell under the Canadian Emergency Department Triage and Acuity Scale (CTAS) IV or V, indicating less urgent or nonurgent cases [14]. It is reasonable to assume that many of these patients could have avoided ED visits by providing appropriate transitional care [15].

To address the needs of patients at a higher risk of postdischarge complications, the concept of continuity of care through digital home monitoring (DHM) emerges as a promising avenue to enhance education, modify behavior, and ultimately achieve improved patient outcomes [9]. With this approach, care teams gain insights into each patient's condition daily or weekly, eliminating the reliance on sporadic office visits, typically occurring only once or twice a year [16]. This continuous and comprehensive view of patient health empowers care teams to make timely adjustments to care plans and proactively engage patients in self-managing care [17]. A virtual care option that extends postdischarge continuity of care offers a viable solution [18-21].

Given the intricacies of providing continuity of care through DHM and the challenges associated with conducting a well-designed randomized controlled trial (RCT) in this context, a pilot study emerges as an essential preliminary step. The primary aim of this pilot study is to assess acceptability, identify logistical requirements, optimize the study design and data collection process, and evaluate readiness for a full-scale trial [22]. Undertaking an RCT that involves continuity of care with a DHM solution is resource-intensive. It raises practical concerns for all stakeholders, including hospital administrators, nurses, clinicians, and patients. Although the primary objective of this pilot study is to examine the feasibility of conducting a comprehensive RCT, this research specifically aims to investigate the feasibility of continuity of care using DHM on postoperative outcomes in patients following thoracic surgery. We hypothesize that continuity of care facilitated by DHM will reduce 30-day ED visits compared to standard care practices.

Methods

Overview

A parallel-group, 2-arm pragmatic pilot feasibility RCT was conducted from September 2022 to January 2023 at the London Health Sciences Centre. Participants were allocated 1:1 to receive continuity of care with DHM or standard of care (control) following the discharge after their thoracic surgical procedures. All participants provided written or electronic informed consent using the REDCap (Research Electronic Data Capture) tool hosted at the London Health Sciences Centre (REDCap e-consent). The analyses and reporting adhered to the CONSORT (Consolidated Guidelines of Reporting Trials) guidelines for pilot trials and the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist [23,24].

To execute the components of the DHM interventions, the health care team was trained from May 2021 to August 2022 using the Plan-Do-Study-Act cycle. Inclusion criteria were patients aged 18 years or older, undergoing a thoracic surgical procedure

(eg, elective anatomic lung resection or any major foregut procedure, such as an esophagectomy), and the surgeon in agreement with patient enrollment. Accredited thoracic surgeons performed all surgeries. Exclusion criteria were patients with unstable disease processes in the postoperative period (eg, postoperative intensive care unit stay) or those with factors that could impact outcome assessment (eg, cognitive impairment, inability to understand English, and limited access to a telephone, computer, or internet services). Patients were also excluded postoperatively if they had intraoperative or immediate postoperative complications requiring an intensive care unit stay.

Upon enrollment, eligible participants were randomized using the simple randomization feature of REDCap. No stratification factors or blocking were applied. The assignment of groups was concealed until the moment of randomization, at which point REDCap automatically allocated participants to the study arms [25]. All consecutive postoperative patients were approached to participate in the study. The randomization occurred on the day of discharge so that in-hospital care was not biased. Due to the pragmatic nature of the trial, patients, surgeons, clinical navigators (CNs), and other health care providers were not masked in the group allocation.

Preoperative, intraoperative, and postoperative patient management followed standard practices and were similar in both groups. A standardized care pathway for postthoracic surgical procedures was implemented for postoperative pain control involving acetaminophen, nonsteroidal anti-inflammatory drugs, hydromorphone, and adjunct medications, such as pregabalin. These were also prescribed on discharge unless otherwise contraindicated. Patients were monitored continuously after surgery while still in the postanesthesia care unit. While on the surgical ward, routine nursing assessments were conducted per the thoracic unit's standard of care. Patients in the control group were discharged home without any monitoring, per the current standard of care. Patients who experienced postdischarge complications were instructed to contact their surgeon's office or visit the hospital ED.

Patients in the DHM group received the same in-hospital care as the control group. In addition, DHM patients signed up for the cloud-based technology platform Vivify Health (Plano, Texas) digital portal with a unique username and password. Through the digital portal, the patient would connect with the CN, who guided the patient through every recovery step. The CN connected, engaged, and educated the patients regarding the recovery pathway. The CN also established clear expectations for patients. Before patients were ready for discharge, patients in the DHM group were given a DHM kit and shown how to use it to maintain continuity of care through the digital care platform. The DHM kit contained a noninvasive blood pressure (NIBP), hemoglobin oxygen saturation (SpO_2), and heart rate (HR) monitor. The data was transferred to a secured digital care platform through the app. DHM patients had access to speak to one of the health care providers at any time of day (CN or virtual care physician). The CN monitored the dashboard from 8 AM to 4 PM After 4 PM, the CN handed over the monitoring dashboard to one of the preassigned physicians (ie, virtual care physicians). Both the CN and physicians were trained in the platform. The health care provider used the digital platform to communicate, engage, and manage patients remotely and efficiently.

Patients measured their vital signs for 2 weeks. The patient also had daily scheduled video calls on days 1-15 after hospital discharge and on an as-needed basis from days 16-30. During the video calls, patients interacted with the CN and responded to symptom questionnaires. The CN organized unscheduled video visits on days without planned virtual visits if they detected changes in patient vital signs or recovery symptom questionnaires requiring follow-up. During virtual visits, the CN discussed any symptoms the patient was experiencing, evaluated their wounds, and obtained a picture if needed. The CN monitored the digital care platform dashboard from the provider side, with an alert for NIBP, HR, SpO_2, wound concerns, home medications, and pain. Alerts were displayed in a color-coded fashion on the dashboard. The CN also monitored the patient's symptoms and identified any changes from the patient's baseline. The CN called a preassigned clinician (ie, the patient's surgeon, a study physician, or a nurse practitioner) if any of the patient's symptoms required medical attention. Physicians could add or modify treatments as needed, and if required, they could have the patient come to an outpatient or ED facility for evaluation or management. Instructions were provided for the patient to call an emergency number (ie, 911) in collaboration and consultation with a physician if appropriate if any symptom indicated immediate distress. The CN and patients were just one button or "mouse click" away from each other, with multiple options to communicate by phone, SMS text message, email, or the virtual care platform (video chat). All these modes of communication were through a secured platform. The CN monitored and intervened by providing patients with advice and next steps if they had health concerns. Self-help educational videos were also available for patients.

This RCT was conducted as a pilot study, with a primary emphasis on assessing feasibility outcomes, which include trial recruitment, protocol adherence, and data collection. We followed the traffic light approach criteria for reporting feasibility outcomes [25-27]. This approach defined (1) feasible (green, 75%-100%) where all feasibility outcomes were met and no protocol modifications were needed; (2) feasible with modification (amber, 50%-75%) where all feasibility outcomes were met or could be met with protocol modifications; and (3) not feasible (red, <50%) where even with protocol modifications, feasibility outcomes could not be met. The clinical outcomes were assessed secondarily to inform the measurement strategy and sample size requirements for a future RCT (ie, by estimating variability, SDs, and prevalence of critical clinical outcomes). Our quadruple health outcome measurement strategy included (1) postoperative outcomes like 30-day ED visits, 30-day RRs, postoperative complications, in-hospital length of stay, 30-day quality of recovery (QoR-40) [28], and postdischarge 30-day opioid consumption; (2) patient-program satisfaction and caregiver satisfaction [29]; (3) health care provider satisfaction; and (4) financial sustainability like cost per case analysis.

Patient-reported outcomes were collected up to 30 days post hospital discharge. Daily data was collected using automatic electronic questionnaires completed digitally and transmitted directly to the REDCap database. Patients also had the option to complete daily questionnaires by video or telephone with a CN. The questionnaires were completed on a smartphone, tablet, or personal computer. Masked assessors verified the data in the REDCap database. Information regarding the 30-day ED visits, 30-day RRs, postoperative complications, postdischarge 30-day opioid consumption, and in-hospital length of stay was obtained from electronic medical records. The patient-program satisfaction survey consisted of 9 questions collected by the research assistant at the end of the 30 days in the DHM group. Patient agreement or satisfaction with statements was recorded on a 5-point scale (from 1=strongly disagree to 5=strongly agree) using a checkmark (✓), with a higher score indicating a higher level of patient agreement or satisfaction. The caregiver survey consists of 17 "Yes" or "No" questions collected by the research assistant at the end of the 30 days in the RPM program in both the DHM and control groups. The satisfaction survey for health care providers comprised 9 questionnaires, addressed at the project's conclusion through the Microsoft Teams survey link and disseminated via electronic mail. Case costing data consisting of the average direct surgical and nonsurgical inpatient costs was obtained for the DHM group and control groups according to the Ontario case costing initiative methodology for 2019-2020 data [30].

The following factors were considered in creating the 5 grades of interventions during the postoperative follow-up using RPM programs: phone calls, video calls, asynchronous messages, self-help educational materials, the amount of time the CN spent addressing the patient's concerns, and escalation to the virtual care physician. The definition of levels of digital health intervention: (1) no intervention and no assessment; (2) no intervention, but the automatic collection of signs, symptoms, and vital signs questionnaires; (3) mild intervention, wherein the CN spends less than 15 minutes with the patient; (4) moderate intervention, characterized by the CN spending 15-30 minutes with the patient; and (5) severe intervention, involving either the CN spending more than 30 minutes with the patient or the situation being escalated to a virtual care physician for further management.

Based on previous data, at least 70 measured participants were required to estimate SD with enough precision for future sample size calculations [31]. We aimed to recruit and obtain outcome data from 80 patients (40 per group), allowing for an attrition rate of approximately 15%. This sample size was also consistent with recommendations regarding the minimum number of participants required to identify feasibility issues [32]. We used "intention-to-treat" analysis. No formal comparison between the study arms was undertaken for outcomes, given that this is a feasibility study. Quantitative secondary outcome measures were summarized descriptively using appropriate summary statistics in the result section and by the trial arm in the tabular column. Continuous variables were reported as mean, standard deviation, and median (range), as appropriate. Categorical variables were reported as counts and percentages. Statistical analyses were performed using GraphPad Prism (GraphPad) software.

Ethical Considerations

This study was formally registered with ClinicalTrials.gov (NCT04340960) and received full board review and approval from the institutional research ethics board at Western University (HSREB 114886). All individual participants involved in the study provided informed consent. Furthermore, appropriate measures were implemented to maintain the confidentiality and anonymity of patient data throughout the research. The study posed no significant risks to the participants, who kept the right to withdraw without facing repercussions regarding their standard of care. Ultimately, no financial compensation was offered to the participants involved in the study.

Results

A total of 91 consecutive patients were considered for inclusion in our study. In total, 80 patients met the inclusion criteria, consented to participate, and were randomized to either the control (n=40) group or the DHM (n=40) group (Figure 1). The 2 groups' patient demographics and clinical characteristics were similar (Table 1).

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram. DHM: digital home monitoring.

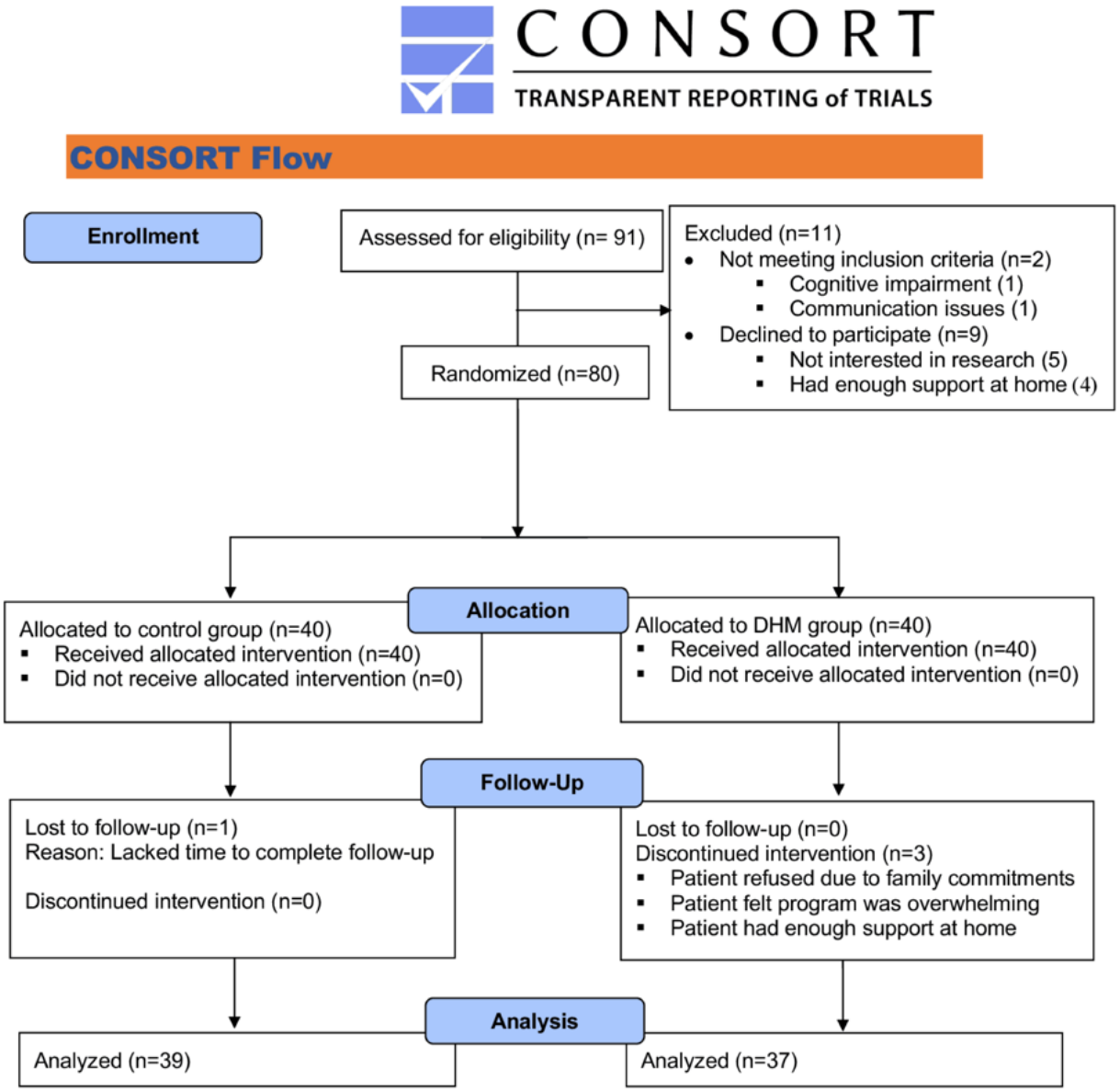


Table 1. Patient demographic and clinical characteristics.

	Control group	DHM[a] group	Total	*P* value
Age (years), mean (SD)	65.5 (14.7)	63.3 (15.0)	64.4 (14.8)	.51
Gender, n (%)				.36
Female	25 (62.5)	20 (50)	45 (56.2)	
Male	15 (37.5)	20 (50)	35 (43.7)	
BMI (kg/m^2), mean (SD)	30.1 (12.3)	27.8 (5.0)	28.9 (9.40)	.28
Outside the London area, n (%)				.48
Yes	23 (57.5)	27 (67.5)	50 (62.5)	
No	17 (42.5)	13 (32.5)	30 (37.5)	
Disease type, n (%)				.99
Primary lung cancer	17 (42.5)	17 (42.5)	34 (42.5)	
Secondary lung cancer	7 (17.5)	7 (17.5)	14 (17.5)	
Others	16 (40)	16 (40)	32 (40)	
PFT, mean (SD) (n)				
FEV1[b]	85.8 (18.2) (16)	88.5 (21.1) (19)	87.3 (19.6) (35)	.69
DLCO[c]	74.6 (19.6) (16)	77.6 (17.1) (18)	76.2 (18.1) (34)	.64
Cancer type, n (%)				.40
Malignant	24 (60)	23 (57.5)	47 (58.7)	
Benign	2 (5)	0 (0)	2 (2.5)	
Others	14 (35)	17 (42.5)	31 (38.7)	
Side of surgery, n (%)				.06
Right	17 (42.5)	8 (20)	25 (31.2)	
Left	10 (25)	18 (45)	28 (35)	
N/A[d]	13 (32.5)	14 (35)	27 (33.7)	
Type of resection, n (%)				.30
Wedge	13 (32.5)	11 (27.5)	24 (30)	
Segmentectomy	4 (10)	1 (2.5)	5 (6.2)	
Lobectomy	8 (20)	12 (30)	20 (25)	
Pneumonectomy	0 (0)	0 (0)	0	
Pleural	3 (7.5)	0 (0)	3 (3.7)	
Mediastinal	1 (2.5)	2 (5)	3 (3.7)	
Foregut procedure	11 (27.5)	14 (35)	25 (31.2)	
Surgical approach, n (%)				.17
Thoracotomy	12 (30)	7 (17.5)	19 (23.7)	
Laparotomy	8 (20)	5 (12.5)	13 (16.2)	
VATS[e]	18 (45)	21 (52.5)	39 (48.7)	
Laparoscopic	2 (5)	7 (17.5)	9 (11.2)	
Staging (pTNM[f]), n (%)				.66
IA/IB	14 (35)	9 (22.5)	23 (28.7)	
IIA/IIB	4 (10)	8 (20)	12 (15)	
IIIA/IIIB	3 (7.5)	2 (5)	5 (6.2)	
IV	2 (5)	1 (2.5)	3 (3.7)	

	Control group	DHM[a] group	Total	*P* value
Metastatic disease	1 (2.5)	2 (5)	3 (3.7)	
N/A	16 (40)	18 (45)	34 (42.5)	
Histology, n (%)				.95
Adenocarcinoma	13 (32.5)	12 (30)	25 (31.2)	
Small cell carcinoma	2 (5)	1 (2.5)	3 (3.7)	
Metastasis	2 (5)	2 (5)	4 (5)	
Others	1 (2.5)	0 (0)	1 (1.2)	
Carcinoid	0 (0)	1 (2.5)	1 (1.2)	
N/A	22 (55)	24 (60)	46 (57.5)	
Smoking history, n (%)				.76
Quit smoking	21 (52.5)	21 (52.5)	42 (52.5)	
Active smokers	5 (12.5)	3 (7.5)	8 (10)	
Nonsmokers	14 (35)	16 (40)	30 (37.5)	

[a]DHM: digital home monitoring.

[b]FEV1: forced expiratory volume at the end of 1 second.

[c]DLCO: diffusing capacity of lung for carbon monoxide.

[d]N/A: not applicable.

[e]VATS: video-assisted thoracoscopy.

[f]pTNM: tumor (T), lymph nodes (N), metastasis (M).

Among the eligible patients who declined enrollment, the most common reason was not being interested in participating in research while receiving care (5.4%), followed by patients having enough support at home for recovery after hospital discharge (4.3%). In total, 3 patients from the DHM group withdrew in the second week after hospital discharge. The first patient withdrew due to family commitments, the second patient felt the program was overwhelming, and the last patient had enough support at home during recovery and decided to withdraw from the program. Only one patient from the control group was lost at the end of the 30-day follow-up period. In total, 76 patients—39 in the control group and 37 in the DHM group—completed the study. Out of 80 caregivers who provided consent for enrollment, only 36 caregivers (16 in the control group and 20 in the DHM group) responded to the caregiver satisfaction questionnaires at the end of the fourth week (47.7%; 95% CI 35.7%-59.1%).

Our study met all green feasibility criteria (Table 2). All 5 thoracic surgeons agreed to have their patients consecutively recruited and adhere to the study protocol. The recruitment rate was 87.9% (95% CI 79.4%-93.8%), and protocol adherence was 96.3% (95% CI 89.4%-99.2%). Data were collected for outcomes in 98.7% (95% CI 92.9%-99.9%) of participants.

Table 2. Feasibility outcomes.

	Not feasible (red)[a]	Feasible with modification (amber)[b]	Feasible (green)[c]	Study results
Trial recruitment	<50%	50%-74%	75%-100%	87.91%
Protocol adherence	<50%	50%-74%	75%-100%	96.25%
Outcome data collection	<50%	50%-74%	75%-100%	98.70%

[a]Not feasible (red) <50%: even with protocol modifications, some feasibility outcomes cannot be met.

[b]Feasible with modification (amber) 50%-75%: all feasibility outcomes are met or can be met with protocol modifications.

[c]Feasible (green) 75%-100%: all feasibility outcomes are met; no protocol modifications are needed.

The mean age of the sample was 64.4 (SD 14.8) years, with 56.2% being female, and the mean BMI was 28.9 (SD 9.4) kg/m^2. Most patients had malignant cancer (58.7%) and primary lung cancer (42.5%). Patients most commonly underwent wedge resection (30%), lobectomy (25%), or foregut procedures (31.2%). The most common surgical approach was video-assisted thoracoscopy (48.7%), followed by thoracotomy (23.7%), and then laparotomy (16.2%).

The mean total length of stay in the hospital was 5.4 (SD 6.6) days (control vs DHM: 6.9, SD 8.8 vs 4.0, SD 2.7), and the incidence of postoperative complications was 33.7% (control vs DHM: 47.5% vs 20%). The total number of ED visits in this sample was 11.2% (control vs DHM: 20.5% vs 2.7%). All these ED visits were unplanned, and the mean time spent in the ED was 894 (SD 1047) minutes (control vs DHM: 1048, SD 1093 vs 130, SD 0). One patient from the DHM group presented to

the ED with testicular pain. Patients from the control group presented with abdominal bloating or distension, wound concerns, dysphagia, or pain crises. The total hospital RR for the sample was 6.5% (control vs DHM: 7.6% vs 5.4%). The unplanned hospital RR was 3.9% (control vs DHM: 7.6% vs 0%), and the planned hospital RR was 2.6% (control vs DHM: 0% vs 5.4%), respectively. The mean 30-day morphine equivalent dose opioid consumption was 92 (SD 94.2) mg (control vs DHM: 74.3, SD 71.9 vs 111.6, SD 110.9), and the mean in-hospital cost per case was CAD $14,729 (US $10,227.96; SD CAD $12,702 [US $8820.40]; control vs DHM: CAD $17,247 [US $11,976.49], SD CAD $15,313 [US $10,633.50] vs CAD $12,145 [US $8433.61], SD CAD $8779 [US $6096.23]; Table 3).

Table 3. Postoperative outcomes.

	Control group	DHM[a] group	Total	*P* value
LOS[b] (days), mean (SD)	6.9 (8.8)	4.0 (2.7)	5.4 (6.6)	.05
Postoperative complications, n (%)	19 (47.5)	8 (20)	27 (33.75)	.01
Unplanned ED[c] visits, n (%)	8 (20.5)	1 (2.7)	9 (11.25)	.02
Planned ED visits, n (%)	0	0	0	—[d]
Time spent in ED (min), mean (SD)	1048 (1093)	130 (0)	894 (1047)	.48
Unplanned RR[e], n (%)	3 (7.6)	0	3 (3.9)	.24
Planned RR, n (%)	0	2 (5.4)	2 (2.6)	.23
Total RR, n (%)	3 (7.6)	2 (5.4)	5 (6.5)	.99
30-Day morphine equivalent dose consumption (mg), mean (SD)	74.3 (71.9)	111.6 (110.9)	92.4 (94.2)	.08
Cost per case (CAD $; CAD $1=US $0.69), mean (SD)	17,247 (15,313)	12,145 (8779)	14,729 (12,702)	.07

[a]DHM: digital home monitoring.

[b]LOS: length of hospital stay.

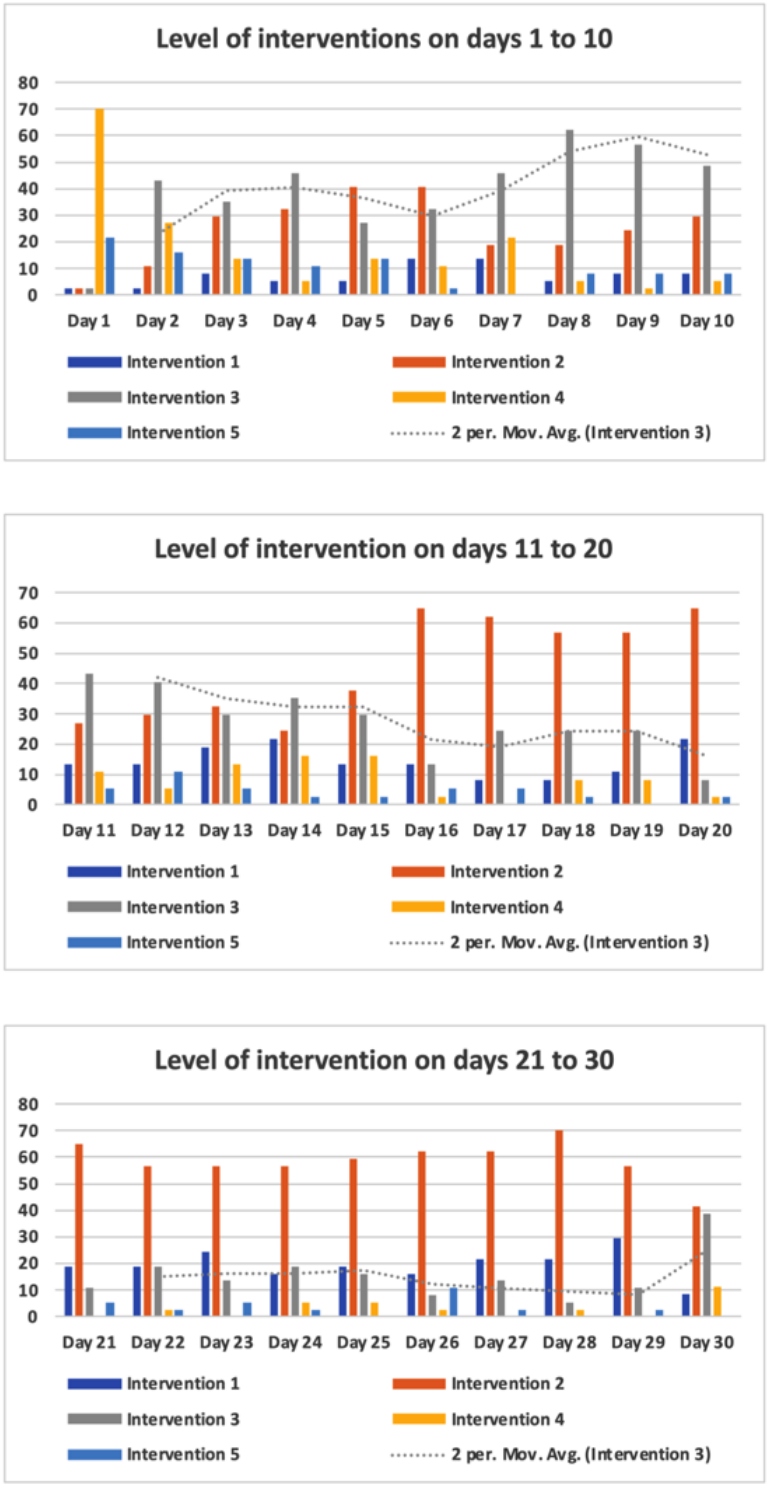
[c]ED: emergency department.

[d]Not applicable.

[e]RR: readmission rate.

The remote monitoring team most often used level 2 or 3 interventions, except for postdischarge day 1, where intervention level 4 was the most common (Figure 2). Comparing interventions over 0-15 days and 16-30 days revealed that level 2 interventions rose significantly from 26.6% to 59.5% (P<.001). In contrast, level 3, 4, and 5 interventions dropped substantially from 38.5% to 16.6%, 15.8% to 3.4%, and 8.6% to 3.2%, respectively (P<.001). The most common issues addressed through the digital platform included pain (23%), surgical wound concerns (11%), shortness of breath (10%), diarrhea (7%), medication management (7%), nausea or vomiting (5%), and dizziness (5%; Multimedia Appendix 1).

Figure 2. Levels of digital health intervention.



At 30 days postoperatively, the mean global QoR-40 score for the sample was 181.9 (SD 5.0). The scores for individual domains included emotional status (39.3, SD 1.4), physical comfort (53.5, SD 0.6), psychological support (33.1, SD 1.7), physical independence (22.7, SD 0.6), and pain (33.1, SD 0.4; Table 4).

Table 4. Quality of recovery.

	Control group (n=39)	DHM[a] group (n=37)	Total (N=76)	*P* value
Global QoR-40[b], mean (SD)	178.3 (3.3)	185.4 (2.6)	181.9 (5.0)	<.001
Emotional status, mean (SD)	38.3 (0.8)	40.4 (0.6)	39.3 (1.4)	<.001
Physical comfort, mean (SD)	53.0 (0.7)	53.9 (0.7)	53.5 (0.6)	<.001
Psychological support, mean (SD)	31.9 (0.5)	34.3 (0.2)	33.1 (1.7)	<.001
Physical independence, mean (SD)	22.2 (0.5)	23.2 (0.4)	22.7 (0.6)	<.001
Pain, mean (SD)	32.7 (0.6)	33.4 (0.5)	33.1 (0.4)	<.001

[a]DHM: digital home monitoring.

[b]QoR-40: 30-day quality of recovery.

Responses from the patient and caregiver satisfaction surveys administered at the end of the fourth week postoperatively were recorded (Multimedia Appendices 2 and 3, respectively). All 37 patients who completed the intervention answered the program satisfaction survey questionnaires (100%; 95% CI 90.5%-100%). More than 95% of patients agreed or strongly agreed that the instructions for setting up the remote monitoring system were easy to understand. All 37 patients in the DHM group agreed or strongly agreed that they felt safe at home and that the CN and physicians responded promptly and efficiently. All patients in the DHM group either agreed or strongly agreed that they would recommend the remote monitoring system program to future patients. Out of 80 caregivers who provided consent for enrollment, only 36 caregivers responded to the caregiver satisfaction questionnaires at the end of the fourth week post hospital discharge (47.7%; 95% CI 35.7%-59.1%). While taking care of the family members at home after the hospital discharge, our sample caregivers reported less burden on family members (8.5%), less interference with personal activities (28.5%), feeling less confined to staying at home (37.1%), and less physical strain (14.2%). However, caregivers reported taking more time off work than initially anticipated (14.2%), employment activities being affected (14.2%), educational activities being affected (8.5%), increased demand on time (31.4%), changes in personal plans (51.4%), and family adjustments (62.8%). Health care providers reported a 100% satisfaction rate (Multimedia Appendix 4).

Discussion

The findings from this trial support the feasibility of conducting a full-scale RCT to compare DHM with the current standard of care after thoracic surgery. The study showed excellent feasibility, achieving a recruitment rate of 87.9%, protocol adherence of 96.3%, and collecting outcome data for 98.7% of participants. These results indicate significant engagement and compliance, reinforcing the study's viability for broader implementation.

The most common barrier to participation among eligible patients in this study was a lack of willingness to participate in research while receiving care (n=5, 5.4%). Other reasons included patients who felt they had enough support at home to recover after hospital discharge (n=4, 4.3%). However, all patients consented to randomization due to the preconception that the care team would connect with them after hospital discharge to aid their recovery. This finding suggests that recruitment for a full-scale trial may be facilitated by addressing implicit biases and emphasizing the importance of continuous connection with the care team to improve postoperative outcomes. Most patients preferred being assigned to the continuity of care with a DHM group rather than the standard care group (70%). In comparison, 25% of the patients did not express any preference.

Using smartphone technology for postoperative follow-up and patient communication has significantly minimized the chances of ED visits and RRs [33,34]. In the United Kingdom, a remote monitoring initiative for 900 colorectal patients reduced costs by 63% while achieving high patient satisfaction [35]. Likewise, a quality improvement study involving 48 thoracic patients with robotic lobectomies found that home monitoring effectively enabled safe early discharges and demonstrated possible economic benefits [36]. Conversely, in an RCT that included 292 postsurgical patients, there was no notable difference between the home monitoring and control groups in ED visits post surgery. Patients in the remote patient monitoring group had an average adherence rate of 86% for daily vital sign logging and 78% for daily question logging [37]. Still, home monitoring was well-received by both patients and physicians, although technological challenges diminished its benefits. Many of these studies relied on automatic data collection methods. Our research yields similar findings but is a prospective RCT focused on thoracic surgical patients. We incorporated more pragmatic inclusion criteria with the caregivers' surveys and used Vivify technology. Our intervention is labor-intensive, differing from other studies, including educational resources, automated questionnaires, vital sign data collection, 2-way communication, and daily CN calls.

This pilot RCT examined the feasibility and clinical impact of continuous DHM on postoperative outcomes in patients undergoing major thoracic surgery. The DHM group had fewer postoperative complications, unplanned ED visits, and unplanned RRs. A potential explanation may be the increased continuity of care and the clinician's ability to monitor a patient's

clinical status to implement necessary interventions before the progression of postoperative complications or ED visits [38-41]. Moreover, the global QoR-40 score and all individual domains were rated higher in the DHM group. This may have resulted from increased patient surveillance and clinician intervention to ensure patients remain on an acceptable path to recovery [42]. However, this trial was not powered or designed with the QoR-40 scores as a primary outcome; thus, these findings must be interpreted cautiously. The satisfaction survey results indicate that patients and health care providers highly value the remote monitoring program. However, caregivers have shown mixed responses. Our findings imply that while patients and providers regard the program positively, further support for caregivers could improve their experience and address the reported increased time demands and schedule adjustments. This can be explored further in the full-scale RCT.

A potential barrier to implementing a DHM system is the difficulty of setting it up and using it by the patient. However, in our study, most patients reported that setup instructions were easy to understand and did not find the system difficult to use. Overall, satisfaction with the program was excellent, and all participants would recommend the remote monitoring system to future patients. Of note, caregivers of patients in the DHM group reported that caregiving affected their personal, educational, and work activities more than the control group. This may be explained by the need to assist the patient in recording vitals and concerns and uploading this information to the digital care platform.

One strength of our study was the diverse patient population regarding gender, age, and BMI. Pathologies such as primary malignancies, secondary malignancies, and nonmalignant diseases were also included. Surgical procedures were diverse, with various types of resections and surgical approaches. The heterogeneity of the study patients indicates that this can be universally implemented in other surgical populations. Patients and their caregivers were adequately trained to record vital signs and upload concerns online, reducing the workload of the home care team. Furthermore, extensive remote patient monitoring was implemented, including HR, NIBP, SpO_2, and daily assessment measurements.

The limitations of this study include the fact that it was not statistically powered to detect postoperative outcome differences. As such, any between-group comparison should be interpreted with caution. Additionally, patients were only followed for 4 weeks postoperatively, so data on the efficacy of continuous DHM on postoperative outcomes beyond this time point remain unknown. Since January 2023, Vivify technology has not been available in Ontario, Canada, and we will be using different technology in our next project to explore these promising results. The potential threats to this feasibility may be reproducibility and scalability associated with the entirely new platform and the maintenance of labor-intensive resource intervention. Further, the cost of the intervention should have been evaluated in this study. Finally, this study was performed at a single center in patients undergoing major thoracic surgery and may need exploration to implement in other surgical populations at different institutions.

In conclusion, the VivifyHealth digital health platform provides a user-friendly interface to extend continuity of care. DHM effectively improved the quality of patients' recovery while decreasing postoperative complications, unplanned ED visits, and hospital RRs. Effective implementation of these platforms may reduce the utilization of scarce health care resources while maintaining excellent patient outcomes and satisfaction. Findings from this pilot trial support the feasibility of conducting a robust full-scale trial to explore these promising results.

Acknowledgments

This study was supported by the Innovation Fund of the Academic Health Sciences Centre—Alternative Funding Plan (AFP) of the Academic Medical Organization of Southwestern Ontario (AMOSO-AFP budget INN20-003), the Department of Anesthesia Internal Research Fund (IRF), and the London Health Sciences Centre Foundation award**.**

Conflicts of Interest

None declared.

Multimedia Appendix 1
Health problems addressed via a digital home monitoring program.
[PNG File , 130 KB - periop_v8i1e58998_app1.png]

Multimedia Appendix 2
Patient/Program Satisfaction Questionnaires.
[DOCX File , 34 KB - periop_v8i1e58998_app2.docx]

Multimedia Appendix 3
Caregivers Satisfaction.
[DOCX File , 34 KB - periop_v8i1e58998_app3.docx]

Multimedia Appendix 4
Healthcare Provider satisfaction.
[DOCX File , 33 KB - periop_v8i1e58998_app4.docx]

Multimedia Appendix 5
CONSORT eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 431 KB - periop_v8i1e58998_app5.pdf]

References

1. Dickinson KJ, Taswell JB, Allen MS, Blackmon SH, Nichols FC, Shen R, et al. Unplanned readmission after lung resection: complete follow-Up in a 1-year cohort with identification of associated risk factors. Ann Thorac Surg 2017;103(4):1084-1091. [doi: 10.1016/j.athoracsur.2016.09.065] [Medline: 27993376]
2. Konstantinidis K, Woodcock-Shaw J, Dinesh P, Brunelli A. Incidence and risk factors for 90-day hospital readmission following video-assisted thoracoscopic anatomical lung resection†. Eur J Cardiothorac Surg 2019;55(4):666-672. [doi: 10.1093/ejcts/ezy345] [Medline: 30364954]
3. Lee AJ, Liu X, Borza T, Qin Y, Li BY, Urish KL, et al. Role of post-acute care on hospital readmission after high-risk surgery. J Surg Res 2019;234:116-122 [FREE Full text] [doi: 10.1016/j.jss.2018.08.053] [Medline: 30527462]
4. Kidane B, Higgins S, Hirpara DH, Kaaki S, Shen YC, Allison F, et al. From emergency department visit to readmission after esophagectomy: analysis of burden and risk factors. Ann Thorac Surg 2021;112(2):379-386. [doi: 10.1016/j.athoracsur.2020.11.020] [Medline: 33310147]
5. Nelson DB, Lapid DJ, Mitchell KG, Correa AM, Hofstetter WL, Mehran RJ, et al. Perioperative outcomes for stage I non-small cell lung cancer: differences between men and women. Ann Thorac Surg 2018;106(5):1499-1503. [doi: 10.1016/j.athoracsur.2018.06.070] [Medline: 30118712]
6. Yang H, Dervin G, Madden S, Beaulé PE, Gagné S, Crossan ML, et al. Postoperative home monitoring after joint replacement: feasibility study. JMIR Perioper Med 2018;1(2):e10168 [FREE Full text] [doi: 10.2196/10168] [Medline: 33401364]
7. Yang H, Dervin G, Madden S, Fayad A, Beaulé P, Gagné S, et al. Postoperative home monitoring after joint replacement: retrospective outcome study comparing cases with matched historical controls. JMIR Perioper Med 2018;1(2):e10169 [FREE Full text] [doi: 10.2196/10169] [Medline: 33401365]
8. van Walraven C, Mamdani M, Fang J, Austin PC. Continuity of care and patient outcomes after hospital discharge. J Gen Intern Med 2004;19(6):624-631 [FREE Full text] [doi: 10.1111/j.1525-1497.2004.30082.x] [Medline: 15209600]
9. Minervini F, Taylor J, Hanna WC, Agzarian J, Hughes K, Pinkney P, et al. Can a mobile app technology reduce emergency department visits and readmissions after lung resection? A prospective cohort study. Can J Surg 2022;65(6):E798-E804 [FREE Full text] [doi: 10.1503/cjs.000122] [Medline: 36418065]
10. Leppin AL, Gionfriddo MR, Kessler M, Brito JP, Mair FS, Gallacher K, et al. Preventing 30-day hospital readmissions: a systematic review and meta-analysis of randomized trials. JAMA Intern Med 2014;174(7):1095-1107 [FREE Full text] [doi: 10.1001/jamainternmed.2014.1608] [Medline: 24820131]
11. Young JM, Butow PN, Walsh J, Durcinoska I, Dobbins TA, Rodwell L, et al. Multicenter randomized trial of centralized nurse-led telephone-based care coordination to improve outcomes after surgical resection for colorectal cancer: the CONNECT intervention. J Clin Oncol 2013;31(28):3585-3591. [doi: 10.1200/jco.2012.48.1036]
12. Shargall Y, Hanna WC, Schneider L, Schieman C, Finley CJ, Tran A, et al. The integrated comprehensive care program: a novel home care initiative after major thoracic surgery. Semin Thorac Cardiovasc Surg 2016;28(2):574-582. [doi: 10.1053/j.semtcvs.2015.12.003] [Medline: 28043480]
13. Fox JP, Suter LG, Wang K, Wang Y, Krumholz HM, Ross JS. Hospital-based, acute care use among patients within 30 days of discharge after coronary artery bypass surgery. Ann Thorac Surg 2013;96(1):96-104 [FREE Full text] [doi: 10.1016/j.athoracsur.2013.03.091] [Medline: 23702228]
14. Book. All-Cause Readmission to Acute CareReturn to the Emergency Department. Ottawa, ON: Canadian Institute for Health Information; 2012.
15. Cleeland CS, Wang XS, Shi Q, Mendoza TR, Wright SL, Berry MD, et al. Automated symptom alerts reduce postoperative symptom severity after cancer surgery: a randomized controlled clinical trial. J Clin Oncol 2011;29(8):994-1000. [doi: 10.1200/jco.2010.29.8315] [Medline: 21282546]
16. Sun V, Dumitra S, Ruel N, Lee B, Melstrom L, Melstrom K, et al. Wireless monitoring program of patient-centered outcomes and recovery before and after major abdominal cancer surgery. JAMA Surg 2017;152(9):852-859 [FREE Full text] [doi: 10.1001/jamasurg.2017.1519] [Medline: 28593266]
17. Cheng X, Yang Y, Shentu Y, Ding Z, Zhou Q, Tan Q, et al. Remote monitoring of patient recovery following lung cancer surgery: a messenger application approach. J Thorac Dis 2021;13(2):1162-1171 [FREE Full text] [doi: 10.21037/jtd-21-27] [Medline: 33717589]

18. Avery KNL, Richards HS, Portal A, Reed T, Harding R, Carter R, et al. Developing a real-time electronic symptom monitoring system for patients after discharge following cancer-related surgery. BMC Cancer 2019;19(1):463 [FREE Full text] [doi: 10.1186/s12885-019-5657-6] [Medline: 31101017]
19. Keng CJS, Goriawala A, Rashid S, Goldstein R, Schmocker S, Easson A, et al. Home to stay: an integrated monitoring system using a mobile app to support patients at home following colorectal surgery. J Patient Exp 2020;7(6):1241-1246 [FREE Full text] [doi: 10.1177/2374373520904194] [Medline: 33457571]
20. Hong TH, Kim MK, Ryu DJ, Park JS, Bae GC, Jeon YS. The reliability of remote patient-reported outcome measures via mobile apps to replace outpatient visits after rotator cuff repair surgery: repetitive test-retest comparison study for 1-year follow-up. J Med Internet Res 2021;23(3):e20989 [FREE Full text] [doi: 10.2196/20989] [Medline: 33646133]
21. Subramani Y, Querney J, Singh P, Zhang Y, Fochesato L, Fatima N, et al. Preoperative anesthesia virtual video consultations in a preadmission clinic: quality improvement study. JMIR Perioper Med 2024;7:e57541 [FREE Full text] [doi: 10.2196/57541] [Medline: 39052992]
22. Kistin C, Silverstein M. Pilot studies: a critical but potentially misused component of interventional research. J Am Med Assoc 2015;314(15):1561-1562 [FREE Full text] [doi: 10.1001/jama.2015.10962] [Medline: 26501530]
23. Eldridge SM, Chan CL, Campbell MJ, Bond CM, Hopewell S, Thabane L, PAFS Consensus Group. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. Pilot Feasibility Stud 2016;2(1):64 [FREE Full text] [doi: 10.1186/s40814-016-0105-8] [Medline: 27965879]
24. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
25. Singh J. Randomization and online databases for clinical trials. J Pharmacol Pharmacother 2014;5(2):173-174. [doi: 10.4103/0976-500x.130155]
26. Thabane L, Lancaster G. A guide to the reporting of protocols of pilot and feasibility trials. Pilot Feasibility Stud 2019;5(1):37 [FREE Full text] [doi: 10.1186/s40814-019-0423-8] [Medline: 30858987]
27. Avery KNL, Williamson PR, Gamble C, O'Connell Francischetto E, Metcalfe C, Davidson P, et al. Informing efficient randomised controlled trials: exploration of challenges in developing progression criteria for internal pilot studies. BMJ Open 2017;7(2):e013537. [doi: 10.1136/bmjopen-2016-013537] [Medline: 28213598]
28. Gornall B, Myles P, Smith C, Burke J, Leslie K, Pereira M, et al. Measurement of quality of recovery using the QoR-40: a quantitative systematic review. Br J Anaesth 2013;111(2):161-169 [FREE Full text] [doi: 10.1093/bja/aet014] [Medline: 23471753]
29. Nagappa M, Querney J, Martin J, John-Baptiste A, Subramani Y, Lanting B, et al. Perioperative satisfaction and health economic questionnaires in patients undergoing an elective hip and knee arthroplasty: a prospective observational cohort study. Anesth Essays Res 2021;15(4):413-438 [FREE Full text] [doi: 10.4103/aer.aer_5_22] [Medline: 35422546]
30. OCC. Ontario Case Costing Initiative. URL: http://www.occp.com/mainPage.htm? [accessed 2015-07-18]
31. Teare MD, Dimairo M, Shephard N, Hayman A, Whitehead A, Walters SJ. Sample size requirements to estimate key design parameters from external pilot randomised controlled trials: a simulation study. Trials 2014;15(1):264 [FREE Full text] [doi: 10.1186/1745-6215-15-264] [Medline: 24993581]
32. Viechtbauer W, Smits L, Kotz D, Budé L, Spigt M, Serroyen J, et al. A simple formula for the calculation of sample size in pilot studies. J Clin Epidemiol 2015;68(11):1375-1379. [doi: 10.1016/j.jclinepi.2015.04.014] [Medline: 26146089]
33. Lu K, Marino NE, Russell D, Singareddy A, Zhang D, Hardi A, et al. Use of short message service and smartphone applications in the management of surgical patients: a systematic review. Telemed J E Health 2018;24(6):406-414. [doi: 10.1089/tmj.2017.0123] [Medline: 29111887]
34. Dawes AJ, Lin AY, Varghese C, Russell MM, Lin AY. Mobile health technology for remote home monitoring after surgery: a meta-analysis. Br J Surg 2021;108(11):1304-1314. [doi: 10.1093/bjs/znab323] [Medline: 34661649]
35. Siddika A, Tolia-Shah D, Pearson TE, Richardson NGB, Ross AHM. Remote surveillance after colorectal cancer surgery: an effective alternative to standard clinic-based follow-up. Colorectal Dis 2015;17(10):870-875. [doi: 10.1111/codi.12970] [Medline: 25851058]
36. Mangiameli G, Bottoni E, Tagliabue A, Giudici VM, Crepaldi A, Testori A, et al. Early hospital discharge on day two post-robotic lobectomy with telehealth home monitoring. J Clin Med 2024;13(20):13 [FREE Full text] [doi: 10.3390/jcm13206268] [Medline: 39458218]
37. Spaulding A, Loomis E, Brennan E, Klein D, Pierson K, Willford R, et al. Postsurgical remote patient monitoring outcomes and perceptions: a mixed-methods assessment. Mayo Clin Proc Innov Qual Outcomes 2022;6(6):574-583. [doi: 10.1016/j.mayocpiqo.2022.09.005] [Medline: 36304524]
38. Costantino ME, Frey B, Hall B, Painter P. The influence of a postdischarge intervention on reducing hospital readmissions in a Medicare population. Popul Health Manag 2013;16(5):310-316 [FREE Full text] [doi: 10.1089/pop.2012.0084] [Medline: 23537154]
39. Hu Y, McMurry TL, Isbell JM, Stukenborg GJ, Kozower BD. Readmission after lung cancer resection is associated with a 6-fold increase in 90-day postoperative mortality. J Thorac Cardiovasc Surg 2014;148(5):2261-2267.e1 [FREE Full text] [doi: 10.1016/j.jtcvs.2014.04.026] [Medline: 24823283]

40. Van Haren RM, Correa AM, Sepesi B, Rice DC, Hofstetter WL, Roth JA, et al. Hospital readmissions after pulmonary resection: post-discharge nursing telephone assessment identifies high risk patients. J Thorac Dis 2020;12(3):184-190 [FREE Full text] [doi: 10.21037/jtd.2020.02.08] [Medline: 32274083]
41. Varela G, Aranda JL, Jiménez MF, Novoa N. Emergency hospital readmission after major lung resection: prevalence and related variables. Eur J Cardiothorac Surg 2004;26(3):494-497. [doi: 10.1016/j.ejcts.2004.05.035] [Medline: 15302041]
42. Gordon WJ, Henderson D, DeSharone A, Fisher HN, Judge J, Levine DM, et al. Remote patient monitoring program for hospital discharged COVID-19 patients. Appl Clin Inform 2020;11(5):792-801 [FREE Full text] [doi: 10.1055/s-0040-1721039] [Medline: 33241547]

Abbreviations

CN: clinical navigator
CONSORT: Consolidated Standards of Reporting Trials
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
CTAS: Canadian Emergency Department Triage and Acuity Scale
DHM: digital home monitoring
ED: emergency department
HR: heart rate
NIBP: noninvasive blood pressure
QoR-40: 30-day quality of recovery
RCT: randomized controlled trial
REDCap: Research Electronic Data Capture
RR: readmission rate
SpO2: hemoglobin oxygen saturation

Edited by R Saffary; submitted 05.04.24; peer-reviewed by J Shiffermiller; comments to author 30.10.24; revised version received 11.11.24; accepted 05.01.25; published 12.02.25.

Please cite as:
Nagappa M, Subramani Y, Yang H, Wood N, Querney J, Fochesato LA, Nguyen D, Fatima N, Martin J, John-Baptiste A, Nayak R, Qiabi M, Inculet R, Fortin D, Malthaner R
Enhancing Quadruple Health Outcomes After Thoracic Surgery: Feasibility Pilot Randomized Controlled Trial Using Digital Home Monitoring
JMIR Perioper Med 2025;8:e58998
URL: https://periop.jmir.org/2025/1/e58998
doi:10.2196/58998
PMID:

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[2,3], PhD; Eva van der Meij[2,3], PhD; Judith A F Huirne[2,4], PhD; Johannes R Anema[1,2], PhD

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

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

[3]Department of Public and Occupational Health, Amsterdam UMC location University of Amsterdam, Amsterdam, The Netherlands

[4]Department of Obstetrics and Gynaecology, Amsterdam UMC location 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.

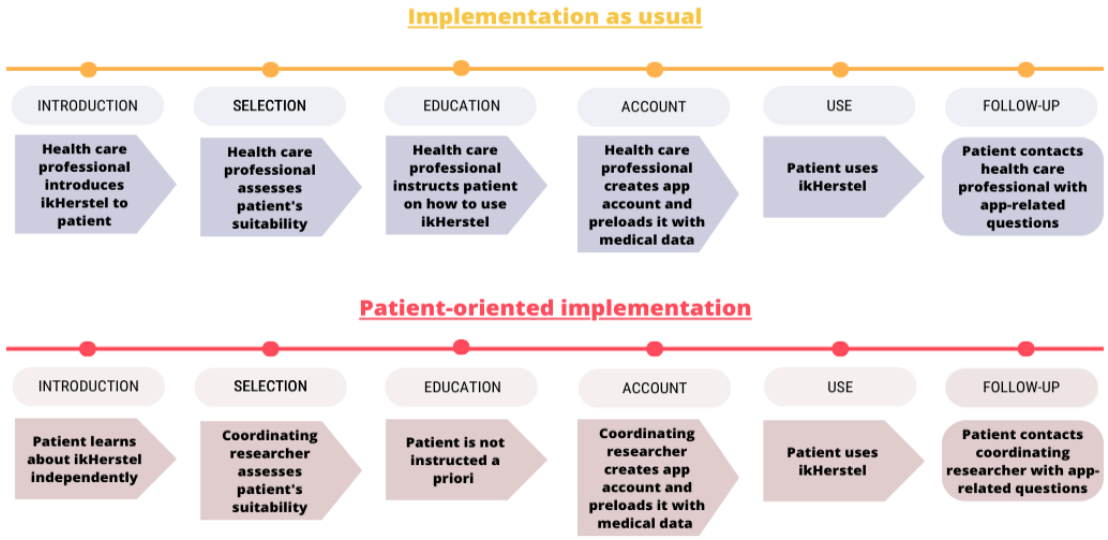


Table 1. Materials used for study recruitment and the frequency of their use, split into hospital-dependent and -independent tools.

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.

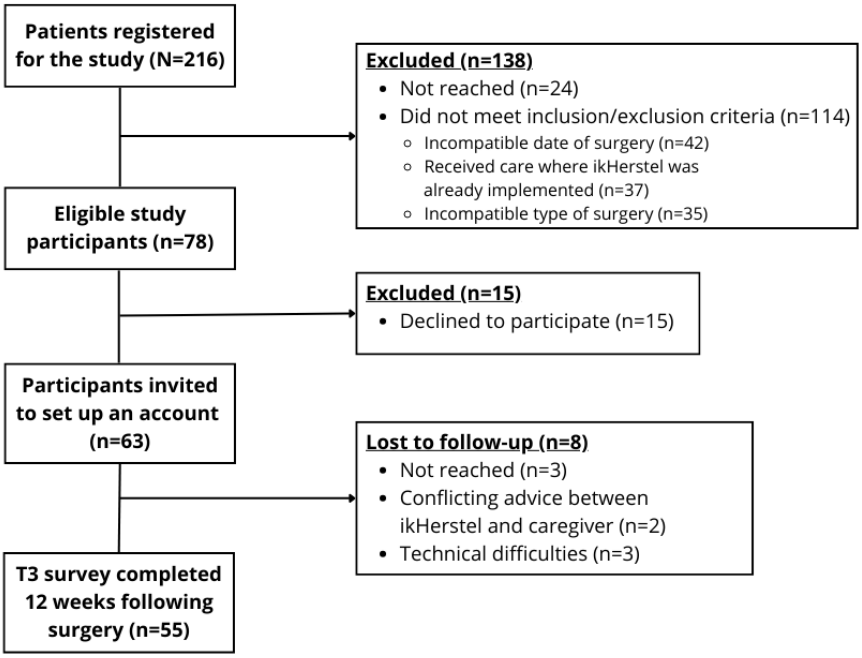


Table 3. Sample characteristics (n=55).

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

Feasibility of a Comprehensive eCoach to Support Patients Undergoing Colorectal Surgery: Longitudinal Observational Study

A Daniëlle Talen[1,2*], MSc, PT; Jobbe P L Leenen[2,3*], PhD, RN; Geert van der Sluis[1,4], PhD, PT; Hilbrand K E Oldenhuis[5], MSc, PhD; Joost M Klaase[6], MD, Prof Dr; Gijsbert A Patijn[2,7], MD, PhD

[1]Research Group Healthy Ageing, Allied Health Care and Nursing, Groningen, Hanze University of Applied Sciences Groningen, Groningen, The Netherlands

[2]Connected Care Center, Isala Hospital, Zwolle, The Netherlands

[3]Research Group IT Innovations in Healthcare, Windesheim University of Applied Sciences, Zwolle, The Netherlands

[4]Department of Health Innovation, Nij Smellinghe Hospital, Drachten, The Netherlands

[5]Research Group Digital Transformation, Hanze University of Applied Sciences, Groningen, The Netherlands

[6]Department of Surgery, University Medical Center Groningen, Groningen, The Netherlands

[7]Department of Surgery, Isala Hospital, Zwolle, The Netherlands

[*]these authors contributed equally

Corresponding Author:
A Daniëlle Talen, MSc, PT
Research Group Healthy Ageing, Allied Health Care and Nursing, Groningen
Hanze University of Applied Sciences Groningen
Petrus Driessenstraat 3
Groningen, 9714 CA
The Netherlands
Phone: 31 0886248103
Email: a.d.talen@isala.nl

Abstract

Background: The mainstay of colorectal cancer care is surgical resection, which carries a significant risk of complications. Efforts to improve outcomes have recently focused on intensive multimodal prehabilitation programs to better prepare patients for surgery, which make the perioperative process even more complex and demanding for patients. Digital applications (eCoaches) seem promising tools to guide patients during their care journey. We developed a comprehensive eCoach to support, guide, and monitor patients undergoing elective colorectal surgery through the perioperative phase of the care pathway.

Objective: The primary aim of this study was to determine its feasibility, in terms of recruitment rate, retention rate, and compliance. Also, usability and patient experience were examined.

Methods: A single-center cohort study was conducted from April to September 2023 in a tertiary teaching hospital in the Netherlands. All elective colorectal surgery patients were offered an eCoach that provided preoperative coaching of the prehabilitation protocol, guidance by giving timely information, and remote monitoring of postoperative recovery and complications. Recruitment and retention rate, as well as compliance for each part of the care pathway, were determined. Secondary, patient-reported usability measured by the Usefulness, Satisfaction, and Ease of Use questionnaire and patient experiences were reported.

Results: The recruitment rate for the eCoach was 74% (49/66). Main reasons for exclusion were digital illiteracy (n=10), not owning a smartphone (n=3), and the expected burden of use being too high (n=2). The retention rate was 80% (37/46). Median preoperative compliance with required actions in the app was 92% (IQR 87-95), and postoperative compliance was 100% (IQR 100-100). Patient-reported usability was good and patient experiences were mostly positive, although several suggestions for improvement were reported.

Conclusions: Our results demonstrate the feasibility of a comprehensive eCoach for guiding and monitoring patients undergoing colorectal surgery encompassing the entire perioperative pathway, including prehabilitation and postdischarge monitoring. Compliance was excellent for all phases of the care pathway and recruitment and retention rates were comparable with rates reported in the literature. The study findings provide valuable insights for the further development of the eCoach and highlight the potential of digital health applications in perioperative support.

(JMIR Perioper Med 2025;8:e67425) doi:10.2196/67425

KEYWORDS

eCoach; telehealth; remote monitoring; home monitoring; virtual; eHealth; colorectal surgery; colorectal cancer; prehabilitation; ERAS; rehabilitation; care pathway; patient journey; feasibility; coaching; mobile phone

Introduction

Background

Colorectal cancer is the third leading cause of cancer deaths worldwide and is mainly diagnosed at an advanced age [1]. The mainstay of colorectal cancer care is surgical resection, which carries a significant risk of complications [2,3]. During hospital admission, enhanced recovery after surgery programs have been adopted widely, resulting in shorter hospital lengths of stay [4,5]. More recently, focus has shifted to optimizing patients preoperatively through multimodal prehabilitation programs, including physical training programs, improving nutritional status, and ameliorating medical comorbidities, thereby reducing postoperative complications [6,7]. After discharge, patients are encouraged to actively rehabilitate to full functional recovery. The entire care pathway from diagnosis to full functional recovery generally takes several months or longer when patients need to receive (neo)adjuvant chemo(radiation) therapy.

For many patients, the perioperative journey can be overwhelming and increasingly complex, as they need to manage a lot of information and perform various tasks at different times [8]. This highlights the need for a broader approach to health care that focuses not just on treating the disease but also on overall well-being, long-term recovery, and self-management. To support patients better, digital tools such as eCoaches are being used more in clinical practice [9]. These tools offer timely information, reminders, and remote monitoring to help patients stay on track and detect complications early. By promoting self-management, eCoaches also reduce the burden on health care systems, which is crucial as resources become more limited [10].

Many health apps for perioperative guidance are available, but the content is often narrow and applied to only one aspect of the care pathway, such as prehabilitation or postoperative monitoring [11-14]. An eCoach for colorectal surgery was reported, but did not include prehabilitation, for which digital coaching can be particularly helpful [15]. Furthermore, clear reporting of feasibility for older surgical patients in real-life clinical practice is often missing [16,17]. A recent study described feasibility of an intervention that combined digital guidance with intensive one-on-one human health coaching, but this is a health professional labor-intensive protocol [18]. More comprehensive digital coaching applications are needed that minimize health care resource usage while optimally informing and engaging patients, ultimately enhancing the quality of care.

A comprehensive eCoach was implemented to guide the patient throughout the perioperative colorectal pathway, providing timely information and monitoring prehabilitation adherence. In addition, immediately after discharge, patients were monitored remotely (vital signs, vomiting, stools, pain, and wound healing with automated alert identification and handling) to allow early detection of postoperative complications and thereby potentially prevent emergency readmissions and improve outcomes. To our knowledge this is the first eCoach for elective colorectal surgery encompassing the entire care pathway, including prehabilitation as well as postdischarge postoperative monitoring. This study explores the feasibility of a digital health application by assessing whether it works as intended in a given context, emphasizing key factors for implementation success while also considering user experience and system demands.

Aim

The primary aim of this study was to determine the feasibility, in terms of recruitment rate, retention rate, and compliance, of a comprehensive eCoach in support of the perioperative care pathway for colorectal surgical patients. The secondary aim was exploring usability, patient experiences and feedback, and evaluating app-induced workload.

Methods

Study Design and Setting

A single-center longitudinal observational study was conducted from April 2023 until September 2023 in a 1200-bed tertiary teaching hospital (Isala, Zwolle) in the Netherlands. Annually, approximately 350 colorectal resections are performed by a team of 5 specialized general surgeons. In April 2023, the eCoach application was implemented into the colorectal surgery pathway at the same time as the implementation of a standardized multimodal prehabilitation program (Fit4Surgery [19]). The STROBE (STrengthening the Reporting of OBservational studies in Epidemiology) guideline for reporting observational studies was followed [20].

Ethical Considerations

The Medical Ethics Committee of the Isala Hospital reviewed the protocol (20230403) and declared that the Medical Research Involving Human Subjects Act (also known by its Dutch abbreviation WMO) did not apply for this study, as the study involves an evaluation of usual care data. The study was conducted in accordance with the Declaration of Helsinki. After onboarding in the eCoach, each patient provided informed consent for use of their personal health information for research purposes in the app.

Participants and Procedures

Patients (older than 18 years) who were preparing for elective colorectal surgery and following the prehabilitation program were included when they were able to communicate in Dutch. The application was integrated into “usual care,” whereby the surgeon explicitly advised patients during their preoperative visit to enroll in the prehabilitation program and to use the app.

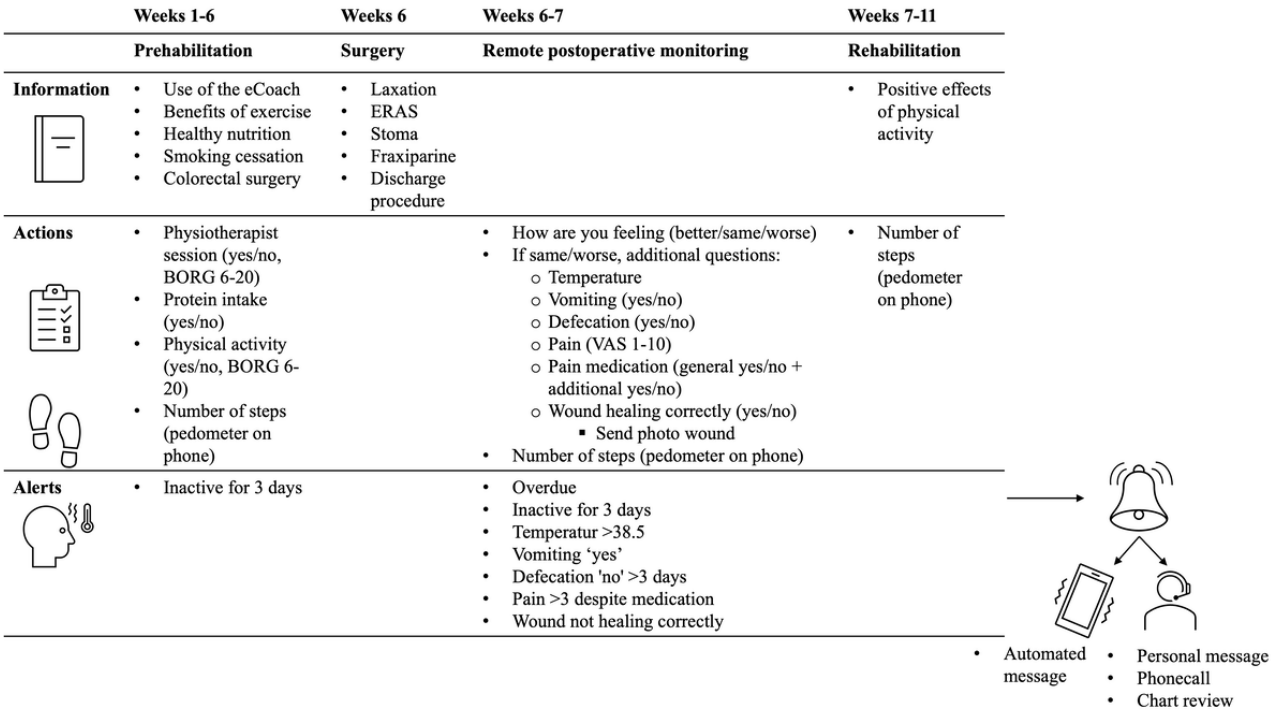
The nurse coordinator checked for eligibility right after the appointment with the surgeon by asking, “do you have a smartphone?” and “are you good at using your smartphone?” Patients were excluded if they were unable to use the app because they did not own a smartphone, did not have web connection, did not possess sufficient digital literacy skills, or had preexistent physical or mental limitations. The onboarding process was completed during an appointment with the case manager (a specially trained nurse who performs the screening and coordination of prehabilitation), who explained the use of the eCoach and evaluated the patient’s ability to use it effectively. Patients who underwent emergency surgery during the care pathway, prior to the planned colorectal resection (eg, due to bowel obstruction), were excluded from the study. When patients received neoadjuvant treatment (radiotherapy or chemoradiotherapy), they were included only after completion and restaging (response evaluation) and definitive acceptance for surgery by the colorectal multidisciplinary team meeting. Health care professionals registered in the electronic patient files if patients were eligible, reasons for nonparticipation, and all usual care data.

Intervention Description

The mobile app eCoach (Luscii Healthtech BV) was developed by health care professionals (clinicians, physiotherapists, dieticians, and nurse practitioners) with expertise on perioperative care in collaboration with the Isala Connected Care team and Luscii Healthtech BV (Multimedia Appendix 1).

Figure 1 illustrates the perioperative care pathway and the integration of the eCoach into this process, including the phases of prehabilitation, surgery, remote postoperative monitoring, and rehabilitation. The eCoach provided tailored information and action prompts specific to each phase. Automated alerts were configured and managed by specialized virtual care nurses at the Isala Virtual Care Center. If required actions were not completed, an automated reminder was sent in the evening. Inactivity for more than 3 consecutive days triggered an alert to the virtual care nurse, who could then take appropriate action, such as sending a personal message, making a phone call, or reviewing the patient’s chart and deciding that no action was necessary. The eCoach acts as a gatekeeper, with all processes being highly standardized and objective. The virtual care nurse reviews the situation when an alert is triggered and determines the appropriate action based on the specific circumstances. This ensures that patient management is consistent and reliable, while allowing for personalized intervention when necessary.

Figure 1. Overview of the eCoach intervention in the colorectal care pathway. BORG: Borg Rating of Perceived Exertion Scale; ERAS: enhanced recovery after surgery protocol; VAS: visual analogue scale.



In the prehabilitation phase (weeks 1-6), the eCoach monitored adherence to the multimodal prehabilitation program and provided timely information on relevant aspects of the care pathway. This involved prompting patients to report whether they attended a physiotherapy session that day and to record their Borg Rating of Perceived Exertion (BORG 6-20) as soon as possible afterward. To support adherence to the nutritional component, the eCoach inquired whether patients had taken their prescribed protein supplement.

During the surgery phase (week 6), the eCoach provided information on key topics such as preparing for surgery, bowel preparation (if applicable), anticoagulant therapy, and discharge procedures.

In the remote postoperative monitoring phase (weeks 6-7), patients completed a daily questionnaire assessing how they felt compared with the previous day (better, the same, or worse). If they reported feeling the same or worse, the eCoach prompted additional questions about body temperature, pain, vomiting, defecation, and wound healing. If responses exceeded set thresholds, an alert was generated, and an automated message advised the patient to contact the hospital. The virtual care nurse checked the alerts and took action when necessary. This included validating the alerts (eg, requesting wound details and photographs and forwarding them to the responsible department) and ensuring that patients followed the app's advice to contact the hospital.

In the rehabilitation phase (weeks 7-11) the eCoach provided information about the positive effect of physical activity in recovery after surgery and monitoring of physical activity. Although the care pathway transitions into long-term cancer follow-up, in this feasibility study, it was considered to end at 30 days post surgery.

Variables and Measurements

Primary Outcomes

The primary outcome of this study was the feasibility of the eCoach app. Feasibility explores whether a digital health system works as intended in a given context and was measured by the recruitment rate, retention rate, and compliance [21]. Recruitment rate was calculated as the proportion of eligible patients, relative to the total elective colorectal surgery patient cohort during the study period. Retention rate was defined as the proportion of patients who completed the use of the eCoach until the end of the eCoach care pathway (30 days after surgery), with reasons for dropout documented. Compliance was defined as the extent at which patients followed the prescribed actions (as shown in Figure 1) within the app, as presented in the intervention description. Since the rates reported in previous literature range between 53% and 95%, we deemed the eCoach feasible when the recruitment rate, retention rate, and compliance were all above 70% [17].

Secondary Outcomes

The secondary outcomes were patient experiences (usability and feedback), app-induced nursing activities, and preliminary effectiveness parameters. A full description of the operationalization of the secondary outcomes can be found in Multimedia Appendix 2.

Usability was evaluated using the Usefulness, Satisfaction, and Ease of Use (USE) questionnaire, which consists of 30 statements rated on a 7-point Likert scale [22]. These statements pertain to 4 key constructs: usefulness, satisfaction, ease of use, and ease of learning regarding the interventions. The questionnaire had been translated into Dutch and used in prior research, with Cronbach a per construct from 0.916 to 0.965 [23,24]. It was gathered using an automatic message in the eCoach, which included a link to the questionnaire.

Feedback on the app's use was collected at the end of the telemonitoring process by the virtual care nurse through a phone call, which was documented in the electronical patient dossier. During the call, patients were asked open-ended questions such as, "How did you experience this process?" and "What improvements would you suggest?" Patient feedback was coded and the themes were categorized into "positive experiences" and "proposed improvements" applying the principles of content analysis [25]. Coding and thematizing was performed by 2 researchers (ADT and JPLL) who discussed differences until consensus was reached. The number of times a theme was mentioned by a patient was reported.

App-induced nursing activities were determined by describing the number of alerts per action item as described in Figure 1 and type of nurse actions that were initiated by alerts of the eCoach.

Preliminary effectiveness parameters consisted of preoperative outcomes after prehabilitation (Steep Ramp Test, 1 repetition maximum tests, and Patient-Generated Subjective Global Assessment Short Form), perioperative functional outcomes (quality of recovery, physical functioning, and quality of life), and postoperative parameters (postoperative complications, length of stay, and time to functional recovery). The Quality of Recovery-15 (QoR-15), Patient-Reported Outcomes Measurement Information System (PROMIS)–Physical Function (PF), and PROMIS-10 questionnaires were administered preoperatively (1 day before surgery) and postoperatively (2 days, 7 days, and 30 days) automatically in the eCoach app. The virtual care nurse made a scheduled call to all patients to remind them about the 30-day questionnaires and asked their feedback on the process. Patient characteristics and postoperative parameters of the study population were gathered (Table 1).

Table 1. Patients' characteristics.

Characteristics	Study population (n=37)
Sex (female), n (%)	17 (46)
Age (year), median (IQR)	65 (60-77)
BMI, median (IQR)	26 (24-32)
ASA[a], n (%)	
I	3 (8)
II	24 (65)
III	9 (24)
IV	1 (3)
CCI[b], median (IQR)	5 (4-6)
Type of surgery (laparoscopic), n (%)	36 (97)
Tumor location, n (%)	
Colon	27 (73)
Rectum	10 (27)
Tumor sort (malignant), n (%)	36 (97)
Surgery procedure, n (%)	
Right hemicolectomy	14 (38)
Left hemicolectomy	5 (14)
Sigmoid resection	8 (22)
LAR[c]	7 (19)
APR[d]	2 (5)
Stoma	1 (3)
Smoking (yes), n (%)	4 (11)
VSAQ[e], median (IQR)	8 (6-10)
HADS[f], median (IQR)	5 (3-8.5)
Hemoglobin (mmol), median (IQR)	
Baseline (n=33)	8.2 (7.0-9.2)
Preoperative (n=17)	7.6 (6.0-9.1)
Complications (yes), n (%)	8 (22)
Clavien-Dindo, n (%)	
I-II	4 (11)
III	3 (8)
IV	1 (3)
Length of stay in days, median (IQR)	4 (3-5)
Time to functional recovery, median (IQR)	1 (0-2)
Care after discharge, n (%)	
Independent	28 (76)
Home care	9 (24)
Rehabilitation center	N/A[g]
Readmissions (yes), n (%)	3 (8)
SRT[h] (W[i]/kg), median (IQR)	2.61 (2.10-3.68) before prehabilitation[j]; 2.82 (2.16-4.10) after prehabilitation[k]

Characteristics	Study population (n=37)
1 RM[l] tests, median (IQR)	
Row	30 (24-35) before prehabilitation; 35 (28-43) after prehabilitation
Chest press	40 (26-56) before prehabilitation; 53 (35-60) after prehabilitation
Leg press	210 (183-323) before prehabilitation; 235 (188-299) after prehabilitation
Lat pulldown	32 (26-38) before prehabilitation; 35 (29-42) after prehabilitation
PG-SGA sf[m], median (IQR)	1 (0.5-4) before prehabilitation; 1 (0-2) after prehabilitation

[a]ASA: American Society of Anesthesiologists.
[b]CCI: Charlson Comorbidity Index.
[c]LAR: low anterior resection.
[d]APR: abdominal perineal resection.
[e]VSAQ: Veteran Specific Activity Questionnaire.
[f]HADS: Hospital Anxiety and Depression Scale.
[g]N/A: not applicable.
[h]SRT: Steep Ramp Test.
[i]W: wattage.
[j]Before prehabilitation: n=34.
[k]After prehabilitation: n=26.
[l]RM: repetition maximum.
[m]PG-SGA sf: Patient-Generated Subjective Global Assessment Short Form.

Statistical Analysis

Formal sample size calculation was challenging given the observational feasibility study design, but a sample in the range of 20-25 is considered adequate for this type of study [26,27]. We determined to include all patients during 3 months. Given the expected number of surgeries (n=85) and dropout rate of previous studies (50%), we expected to include 40 patients.

Descriptive statistics were used to evaluate patient demographics and to assess the feasibility. Continuous data were checked for normality by a Shapiro-Wilk test and visually by a histogram. Based on normality, median and IQR or mean and SD were presented. For categorical data, frequencies and percentages were calculated. All data were analyzed using SPSS Statistics (version 24; IBM Corp) for Windows. The answers on the open-ended questions were coded and categorized by 2 researchers (ADT and JPLL) by content analysis based with predefined categories: positive experiences and proposed improvements. Categories were quantified.

Results

Overview

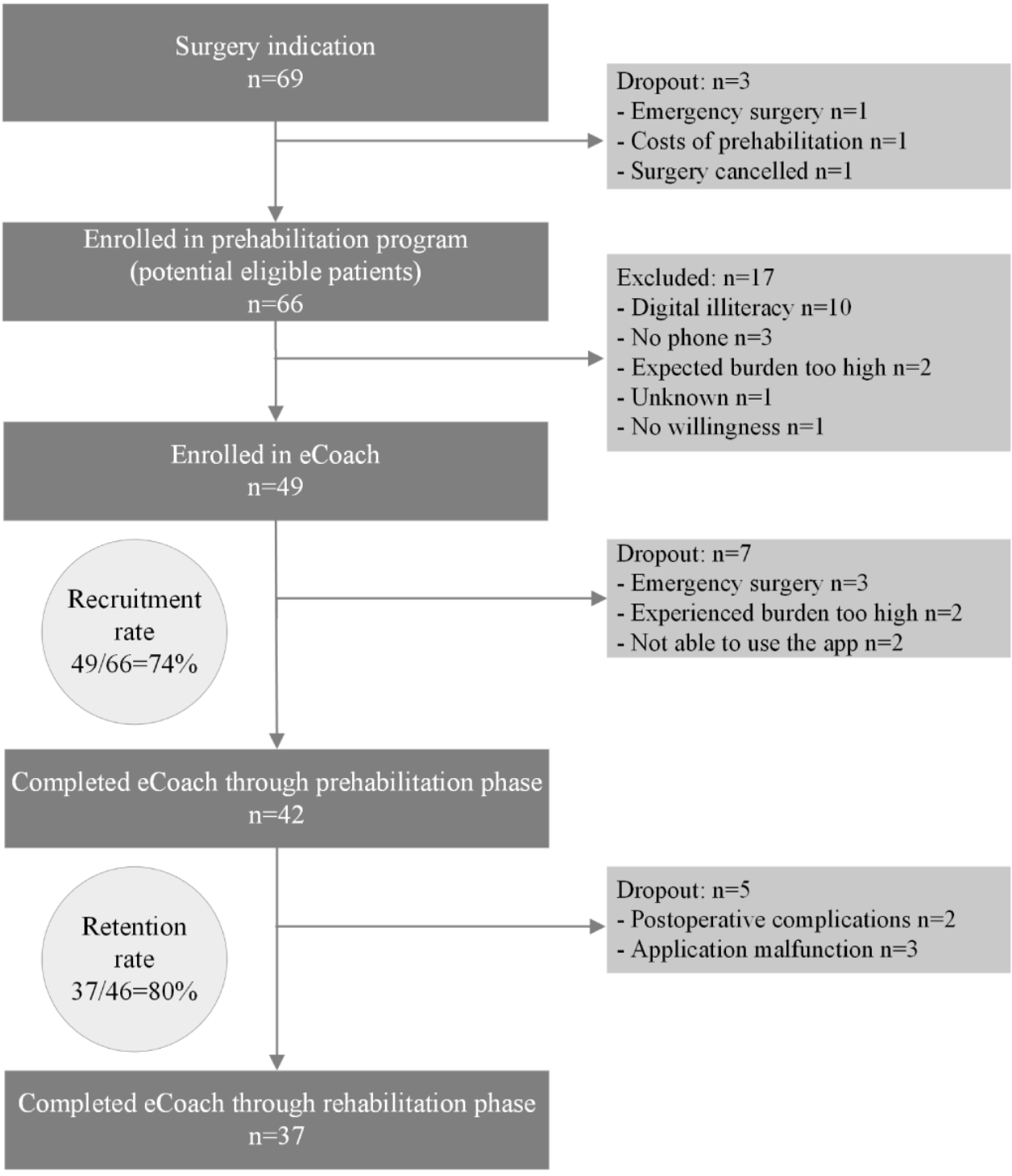
A total of 37 patients completed the study, of which the patient characteristics are shown in Table 1. Median age was 65 (IQR 60-77) years and median Charlson Comorbidity Index score was 5 (IQR 4-6). Four patients experienced minor complications (Clavien Dindo I-II), such as issues with ileostomy production and atrial fibrillation. The other 4 patients had major complications (Clavien Dindo III-IV), including abscess, anastomotic leakage, and systemic inflammatory response syndrome, with 1 patient requiring intensive care unit admission.

Primary Outcomes

Recruitment Rate and Retention Rate

During the study period 66 patients were eligible for the colorectal surgery pathway. Of the 66 included patients, 49 enrolled in the eCoach, resulting in a recruitment rate of 74% (49/66; Figure 2). Main reasons for exclusion were digital illiteracy (n=10), not having a phone (n=3), and expected extra burden of the app being too high (n=2). Of the 49 enrolled patients, 4 (92%) dropped out preoperatively during the prehabilitation phase and 5 (88%) postoperatively during the close monitoring or rehabilitation phase, resulting in a retention rate of 80%. Two patients dropped out due to the significant burden imposed by postoperative complications, leading them to discontinue using the eCoach. The 3 patients who underwent emergency surgery were excluded from the calculations of recruitment and retention rates.

Figure 2. Flowchart of study population, including recruitment rate and retention rate.



Compliance

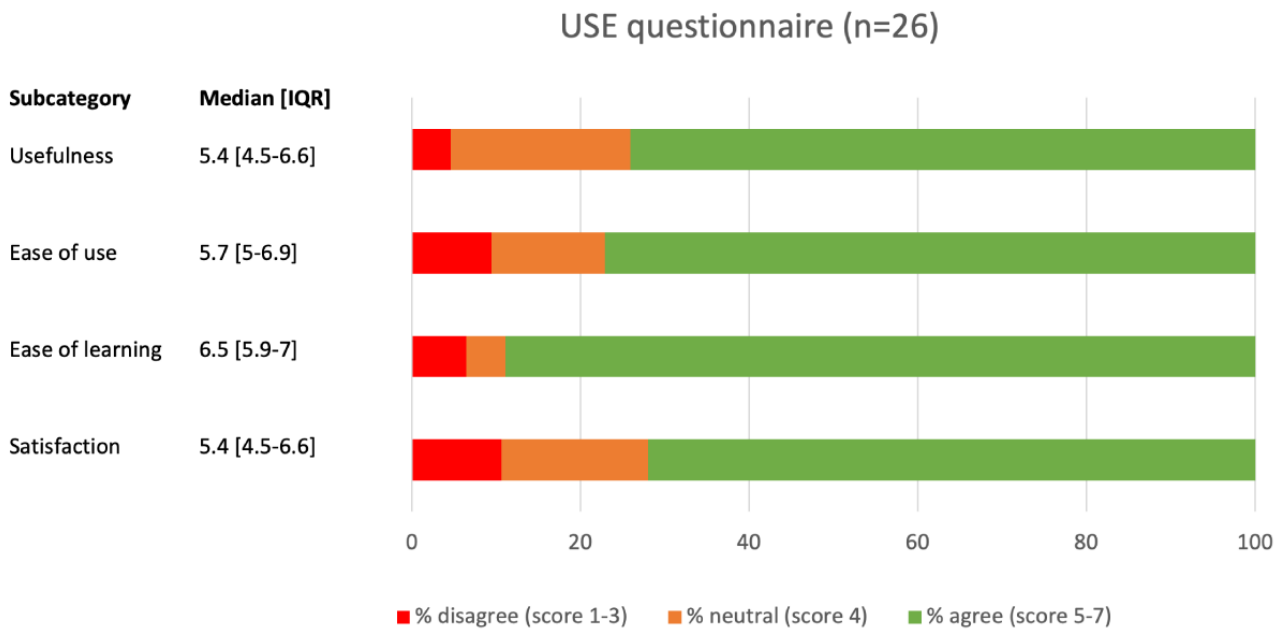
Median compliance was 95% (IQR 82%-96%), preoperative compliance was 92% (IQR 87%-95%), and postoperative compliance was 100% (IQR 100%-100%). Preoperative compliance was highest with 98% (IQR 90%-100%) with "physiotherapist session." Compliance with the number of steps was lowest with 86% (IQR 72%-93%). Compliance with "protein intake" was 90% (IQR 84%-97%), on which patients reported "yes" 97% of the time. Of the 37 patients, 34 patients responded. The median compliance to the postoperative questions "Well-being compared to yesterday" was 100% (IQR 100%-100%), and patients reported feeling better 60% of the times. Compliance for the additional questions was 100% (IQR 100%-100%), where "Wound healing correctly" resulted in a negative response most of the times. A comprehensive presentation of compliance with various components of the eCoach is shown in Multimedia Appendix 3.

Secondary Outcomes

Patient Experiences: Usability

Twenty-six patients (response rate: 70%) completed the USE questionnaire (Figure 3) at day 30 postoperatively. Median scores for usefulness, ease of use, ease of learning, and satisfaction were 5.4, 5.7, 6.5, and 5.4, respectively, on a 1-7 Likert scale, all of which are considered good outcomes. Scores and IQRs to individual questions and categories are described in detail in Multimedia Appendix 4.

Figure 3. Usefulness, Satisfaction, and Ease of Use questionnaire. USE: Usefulness, Satisfaction, and Ease of Use.



Patient Experiences: Content Analysis

In total, 89% (33/37) of patients answered the questions about their experiences with the eCoach. Forty-eight positive experiences were reported. One patient said, “I especially valued the motivation to stay physically active. I feel like this made me healthier and stronger.” Other patients called it “a good incentive,” or “a helpful reminder for the protein intake.” Twenty-four patients reported 41 proposed improvements. These areas of improvement were diverse, but rigidity of the app was most frequently mentioned (Textbox 1 and Multimedia Appendix 5).

Textbox 1. Content analysis of experiences reported by patients.

Positive experiences (n=48)

- General positive experiences (n=18)
- Providing support and engagement (n=15)
- Informative (n=7)
- Stimulating motivation and incentives (n=4)
- Mental support (n=2)
- Continuous connection (n=2)

Proposed improvements (n=41)

- Limited usability (n=9)
- Rigidity of the app (n=14)
- Problems with the pedometer (n=6)
- Length of postoperative monitoring was unclear or insufficient (n=6)
- Missed features in the app (n=3)
- Engagement difficulties and mental burden (n=3)

App-Induced Nursing Activities

Out of 1752 preoperative alerts, 99.9% (n=35) were processed automatically by the eCoach, with only 2 alerts needing manual interventions for protein intake. Of the 222 alerts for protein intake across 30 patients, 99% (n=29) were processed automatically by the system. Two alerts required manual interference, so 2 messages were sent by the nurse to remind patients about their protein intake. The number of postoperative alerts was 126 (n=10), of which 43% (54/126) were processed automatically. The remaining alerts led to 21 phone calls and 19 messages in the app. The alerts for "Wound healing correctly" resulted in the most alerts with actions necessary. A detailed summary of nursing activities induced by the app is shown in Multimedia Appendix 6.

Preliminary Effectiveness Parameters

The median physical fitness was preoperatively 2.61 (IQR 2.10-3.68) W/kg and postoperatively 2.82 (IQR 2.16-4.10) W/kg on the Steep Ramp Test (Table 1). Eight patients developed complications after surgery (8/37, 22%) and 4 of them severe (Clavien Dindo III or IV) (4/37, 11%). Median length of stay was 4 days (IQR 3-5) and median time to functional recovery was 1 day (IQR 0-2).

Quality of recovery at 30 days postsurgery was rated comparable with preoperative scores, whereas quality of life and physical functioning at 30 days were not completed back to preoperative levels (Table 2).

Table 2. Preliminary effectiveness parameters.

	–2 days (n=22)	+1 day (n=23)	+7 days (n=24)	+30 days (n=26)
Quality of recovery score, median (IQR)	134.5 (105-144)	112.0 (83-120)	119.5 (102-133)	135 (115-143)
PROMIS[a]-10 (quality of life) score, median (IQR)	28 (25-34)	26 (23-29)	N/A[b]	26 (25-32)
PROMIS of Physical Functioning score, median (IQR)	38.5 (23.8-40)	13 (11-16)	N/A	28.5 (20-34)

[a]PROMIS: Patient-Reported Outcomes Measurement Information System.

[b]N/A: not applicable.

Discussion

Principal Findings

Our results demonstrate the feasibility of a comprehensive eCoach that was developed for elective colorectal surgery patients incorporating all phases of the care pathway, including prehabilitation, enhanced recovery after surgery components, and postoperative monitoring. Recruitment (49/66, 74%) and retention (37/46, 80%) rates were comparable with rates reported in the literature, whereas the compliance (overall 95%) was excellent. Patient-reported usability was good, and patients not only reported to value the eCoach as a beneficial addition to the patient journey but also reported some areas of improvement that need to be addressed in future iterations of the eCoach.

Recruitment and Retention

We found that a significant number of eligible patients were unable to use the eCoach. Nineteen patients (19/66, 29%) were excluded at baseline due to digital illiteracy, not owning a smartphone, or finding the eCoach mentally burdensome. Only 1 patient was excluded due to unwillingness to participate. Given the study population of unselected patients with colorectal cancer, including a significant proportion of older adult and frail patients, this finding is, however, not unexpected and in line with recruitment rates reported in the literature [11,18,28-31]. Older age groups are known to have lower digital proficiency and lower smartphone ownership, and using a digital application may provide a high perceived burden for frail older adult patients, who are facing the challenges of a recent cancer diagnosis and an upcoming high-risk surgery [17]. Although our study did not quantify frailty, the reasons for nonparticipation, such as digital illiteracy and lack of smartphone ownership, suggest that excluded patients were more likely to be older adults and vulnerable. This aligns with findings from a digital prehabilitation study, which reported that patients with insufficient digital skills were older and had a more unfavorable risk profile [32].

Retention was comparable or slightly better than rates reported in the literature. Dropouts were disease related, technical, or due to a perceived heavy burden of using the eCoach. The reported technical issues (malfunction) had not been encountered during initial beta testing and were promptly solved by the development team. When dropping out in the preoperative phase due to emergency surgery (eg, bowel obstruction), these patients could no longer participate for external reasons and were thus excluded from the retention assessment.

Our findings show that there is a subset of patients unable to participate or dropping out for various reasons, confirmed by previous studies. Thus, in clinical practice we cannot relying solely on digital coaching. Hybrid approaches including nondigital and personalized coaching and guidance will remain necessary in order to reach all patients. Furthermore, these results highlight the need for more inclusive design in health care technology, ensuring that the development process considers vulnerable groups and digital illiteracy.

Compliance

Overall compliance (adherence) was excellent in our study. It was slightly lower during the preoperative phase (92%) than postoperatively (100%), probably because there were more preoperative actions to comply with. Median compliance for monitoring physiotherapy visits, physical activity, and protein intake varied between 90% and 98%, where in other studies compliance ranged from 53% to 86% [17]. Possible reasons for

the high compliance rates are the user-friendly app design, the seamless integration into routine care, the onboarding meeting with a dedicated case manager, and the explicit encouragement to enroll and adhere from the surgeons. Our results underscore the high potential value of digital applications for encouraging patient engagement and self-management, which may help improve quality of care and reduce the number of unplanned patient-provider contacts.

Usability and Patient Satisfaction

Most of the patients were positive about the guidance by the eCoach, as shown in the high median scores of the USE questionnaire and in the qualitative feedback. Patient feedback has shown that the eCoach also provides implicit psychological support, helping them feel more connected, confident, and mentally prepared for surgery. The qualitative evaluation of a comparable app reported similar responses [33]. Assistance by trained staff is known to increase perceived usability, which might have contributed to our favorable outcomes [17]. Some areas of improvement were reported, including missing features and lack of personalization, which will need to be addressed in future iterations of the eCoach. To further improve patient experiences in real-life practice, the suggested areas of improvement in the feedback should be addressed. Some of these are technical issues, such as links not working correctly, the inability to fill in missed actions a day later, or problems with the pedometer. The problems with the pedometer seem more consistent, since the compliance is structurally lower than the other preoperative actions and the number of alerts, which were all automatically processed, was large. One explanation maybe that some patients did not have a pedometer installed on their smartphone, so often just filled in an estimated number of steps. Furthermore, patients reported inadequate personalization of the eCoach. For instance, patients who were unable to walk (but able to use a bike trainer) felt like the eCoach was not always fitting to their personal situation.

Strengths and Limitations

The key strength of this study is the comprehensive evaluation of the eCoach in real-life clinical practice, covering the entire perioperative pathway for colorectal surgery patients. This practical implementation allows for efficient assessment of feasibility, addressing both patient needs and implementation challenges early on, with high generalizability within the standardized Dutch health care system. The eCoach platform is commercially available and has been adopted by several centers in the Netherlands. In our center, it has become the standard of care and expanded to include complex surgical patients. To support broader implementation, we are committed to training virtual care nurses, sharing our experiences, and establishing virtual care centers. Ensuring that the necessary infrastructure and expertise are in place is essential, and we are actively working toward this to facilitate the expansion of virtual care services [34].

The results of this observational feasibility study have to be interpreted in the light of some limitations, including its single-center design in which the intervention was accessible only to Dutch-speaking people with a web-connected device, and the relatively small cohort. Another limitation of this study is that we included the first group of patients after implementation of the eCoach, potentially resulting in the technical errors experienced by patients. Although these errors were readily addressed during the initial phase of the study, 3 patients had dropped out as a result. It is important to note that these technical issues were part of the initial learning curve and are expected to be minimized in future iterations of the eCoach, ensuring a smoother experience for subsequent patient groups. Furthermore, more alerts were generated initially as we opted to err on the side of caution to ensure patient safety.

Future Directions

Further studies in larger cohorts are needed to assess the potential role of an eCoach in improving clinical effectiveness and cost-efficiency, such as its impact on readmission rates and length of stay, by comparing it with a matched historical control group or randomization [34]. eCoaches may help reduce the burden on the health care system by promoting self-management and compliance and thereby reduce the number of unplanned patient-health care provider contacts. The integration of eCoaches into complex care pathways facilitates comprehensive health management, including approaches that extend beyond traditional disease treatment. Combining the eCoach with an objective measurement device, such as an accelerometer or a continuous vital signs monitoring device, may help reliably measure physical activity and assess time to full recovery [35]. The value of the reported additional secondary end points (QoR, PROMIS-10, and PROMIS-PF) measuring preliminary effectiveness and patient-reported functional parameters were of limited value for this study but will be valuable in future follow-up studies.

As recruitment in virtual eCoach applications will remain suboptimal in older adult or frail patients, studies are needed to develop protocols to better triage patients at baseline to select who are eligible and suitable for inclusion. The complexity of health needs and potential cognitive or physical limitations in older adult, frail, and high-risk patients underscore the need for alternative methods of perioperative support. One way to improve recruitment, retention, and intervention efficacy is designing more personalized and tailored digital health applications [36]. Future iterations of the eCoach may facilitate individual exercise mode by personal choice and tailored communication to the individual level of health literacy and education. Adding a web-based interface or the ability to add a caregiver could reduce the technological barrier for some patients.

Current prehabilitation protocols for colorectal surgery include frequent physical training sessions by physiotherapist. Given the excellent compliance of the eCoach, eligible lower-risk patients may well follow an unsupervised virtual prehabilitation program by using the eCoach. A study to determine the value of unsupervised virtual prehabilitation is planned to start at our institution.

Conclusions

Our results demonstrate the feasibility of using a comprehensive eCoach for guidance and monitoring of elective colorectal surgery patients through all phases of the care pathway.

Compliance was excellent and recruitment and retention rates were comparable with rates reported in the literature. Patient-reported usability was good, and patients reported to value the eCoach as a beneficial addition to the patient journey. The study findings provide valuable insights for the further development of the eCoach and highlight the potential of digital health applications in perioperative support.

Acknowledgments

The authors would like to thank all patients, the Isala prehabilitation case manager, the nurses at the medical coordination office, the colorectal surgery team, and the taskforce prehabilitation for participation and support. The study was funded by the Isala Innovation & Science Funds (grant INNO2023).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Examples of the eCoach interface, showing daily actions for patients on the left and a small part of the available information on the right.
[PNG File , 286 KB - periop_v8i1e67425_app1.png]

Multimedia Appendix 2
Description of secondary outcomes.
[DOCX File , 19 KB - periop_v8i1e67425_app2.docx]

Multimedia Appendix 3
Compliance of the different items of the eCoach.
[DOCX File , 17 KB - periop_v8i1e67425_app3.docx]

Multimedia Appendix 4
USE outcomes.
[DOCX File , 22 KB - periop_v8i1e67425_app4.docx]

Multimedia Appendix 5
Quotes of patient experiences.
[DOCX File , 26 KB - periop_v8i1e67425_app5.docx]

Multimedia Appendix 6
App-induced nursing activities.
[DOCX File , 18 KB - periop_v8i1e67425_app6.docx]

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021;71(3):209-249 [FREE Full text] [doi: 10.3322/caac.21660] [Medline: 33538338]
2. Sharp SP, Malizia R, Skancke M, Arsoniadis EG, Ata A, Stain SC, et al. A NSQIP analysis of trends in surgical outcomes for rectal cancer: what can we improve upon? Am J Surg 2020;220(2):401-407. [doi: 10.1016/j.amjsurg.2020.01.004] [Medline: 31964524]
3. Dharap SB, Barbaniya P, Navgale S. Incidence and risk factors of postoperative complications in general surgery patients. Cureus 2022;14(11):e30975 [FREE Full text] [doi: 10.7759/cureus.30975] [Medline: 36465229]
4. Gustafsson UO, Scott MJ, Hubner M, Nygren J, Demartines N, Francis N, et al. Guidelines for perioperative care in elective colorectal surgery: enhanced recovery after surgery (ERAS) society recommendations: 2018. World J Surg 2019;43(3):659-695. [doi: 10.1007/s00268-018-4844-y] [Medline: 30426190]
5. Lassen K, Soop M, Nygren J, Cox PBW, Hendry PO, Spies C, Enhanced Recovery After Surgery (ERAS) Group. Consensus review of optimal perioperative care in colorectal surgery: enhanced recovery after surgery (ERAS) group recommendations. Arch Surg 2009;144(10):961-969. [doi: 10.1001/archsurg.2009.170] [Medline: 19841366]

6. Wilson RJT, Davies S, Yates D, Redman J, Stone M. Impaired functional capacity is associated with all-cause mortality after major elective intra-abdominal surgery. Br J Anaesth 2010;105(3):297-303 [FREE Full text] [doi: 10.1093/bja/aeq128] [Medline: 20573634]
7. Molenaar CJL, Minnella EM, Coca-Martinez M, Ten Cate DWG, Regis M, Awasthi R, PREHAB Study Group. Effect of multimodal prehabilitation on reducing postoperative complications and enhancing functional capacity following colorectal cancer surgery: the PREHAB randomized clinical trial. JAMA Surg 2023;158(6):572-581 [FREE Full text] [doi: 10.1001/jamasurg.2023.0198] [Medline: 36988937]
8. Chan RJ, Milch VE, Crawford-Williams F, Agbejule OA, Joseph R, Johal J, et al. Patient navigation across the cancer care continuum: an overview of systematic reviews and emerging literature. CA Cancer J Clin 2023;73(6):565-589 [FREE Full text] [doi: 10.3322/caac.21788] [Medline: 37358040]
9. Tsiouris KM, Tsakanikas VD, Gatsios D, Fotiadis DI. A review of virtual coaching systems in healthcare: closing the loop with real-time feedback. Front Digit Health 2020;2:567502 [FREE Full text] [doi: 10.3389/fdgth.2020.567502] [Medline: 34713040]
10. Aapro M, Bossi P, Dasari A, Fallowfield L, Gascón P, Geller M, et al. Digital health for optimal supportive care in oncology: benefits, limits, and future perspectives. Support Care Cancer 2020;28(10):4589-4612 [FREE Full text] [doi: 10.1007/s00520-020-05539-1] [Medline: 32533435]
11. Scheper H, Derogee R, Mahdad R, van der Wal R, Nelissen RGHH, Visser L, et al. A mobile app for postoperative wound care after arthroplasty: ease of use and perceived usefulness. Int J Med Inform 2019;129:75-80. [doi: 10.1016/j.ijmedinf.2019.05.010] [Medline: 31445292]
12. van der Linden MJW, Nahar van Venrooij LMW, Verdaasdonk EGG. Personal devices to monitor physical activity and nutritional intake after colorectal cancer surgery: feasibility study. JMIR Perioper Med 2022;5(1):e40352 [FREE Full text] [doi: 10.2196/40352] [Medline: 36512385]
13. Lambert G, Drummond K, Tahasildar B, Carli F. Virtual prehabilitation in patients with cancer undergoing surgery during the COVID-19 pandemic: protocol for a prospective feasibility study. JMIR Res Protoc 2022;11(5):e29936 [FREE Full text] [doi: 10.2196/29936] [Medline: 35522464]
14. Yi J, Yoon JY, Won CW, Kim M, Lee KS. The roles of health literacy and social support in the association between smartphone ownership and frailty in older adults: a moderated mediation model. BMC Public Health 2024;24(1):1064 [FREE Full text] [doi: 10.1186/s12889-024-18163-z] [Medline: 38632509]
15. Smithson M, McLeod MC, Theiss L, Shao C, Kennedy G, Hollis R, et al. Ileostomy patients using patient engagement technology experience decreased length of stay. J Gastrointest Surg 2022;26(3):635-642 [FREE Full text] [doi: 10.1007/s11605-021-05158-z] [Medline: 34618324]
16. Maramba I, Chatterjee A, Newman C. Methods of usability testing in the development of eHealth applications: a scoping review. Int J Med Inform 2019;126:95-104. [doi: 10.1016/j.ijmedinf.2019.03.018] [Medline: 31029270]
17. Jonker LT, Haveman ME, de Bock GH, van Leeuwen BL, Lahr MMH. Feasibility of perioperative eHealth interventions for older surgical patients: a systematic review. J Am Med Dir Assoc 2020;21(12):1844-1851.e2 [FREE Full text] [doi: 10.1016/j.jamda.2020.05.035] [Medline: 32694000]
18. Esper SA, Holder-Murray J, Meister KA, Lin HS, Hamilton DK, Groff YJ, et al. A novel digital health platform with health coaches to optimize surgical patients: feasibility study at a large academic health system. JMIR Perioper Med 2024;7:e52125 [FREE Full text] [doi: 10.2196/52125] [Medline: 38573737]
19. Sabajo CR, Dieleman JP, Dekker JW, Heuvel B, Klaas JM, Slooter GD. Multimodal prehabilitation for colorectal cancer patients: study protocol of a nationwide multicentre study with uniform prehabilitation protocols. Res Square 2024. [doi: 10.21203/rs.3.rs-4000534/v1]
20. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. BMJ 2007;335(7624):806-808 [FREE Full text] [doi: 10.1136/bmj.39335.541782.AD] [Medline: 17947786]
21. Johns Hopkins University. Monitoring and evaluating digital health interventions: a practical guide to conducting research and assessment. 2016. URL: https://iris.who.int/bitstream/handle/10665/252183/9789241511766-eng.pdf?sequence=1&isAllowed=y [accessed 2025-01-18]
22. Lund AM. Measuring usability with the USE questionnaire. Usability Interface 2001;8(2):3-6.
23. Leenen JPL, Dijkman EM, van Dijk JD, van Westreenen HL, Kalkman C, Schoonhoven L, et al. Feasibility of continuous monitoring of vital signs in surgical patients on a general ward: an observational cohort study. BMJ Open 2021;11(2):e042735 [FREE Full text] [doi: 10.1136/bmjopen-2020-042735] [Medline: 33597138]
24. Gao M, Kortum P, Oswald F. Psychometric evaluation of the USE (Usefulness, Satisfaction, and Ease of use) questionnaire for reliability and validity. Proc Hum Factors Ergon Soc Annu Meet 2018;62(1):1414-1418. [doi: 10.1177/1541931218621322]
25. Bengtsson M. How to plan and perform a qualitative study using content analysis. NursingPlus Open 2016;2:8-14. [doi: 10.1016/j.npls.2016.01.001]

26. Totton N, Lin J, Julious S, Chowdhury M, Brand A. A review of sample sizes for UK pilot and feasibility studies on the ISRCTN registry from 2013 to 2020. Pilot Feasibility Stud 2023;9(1):188 [FREE Full text] [doi: 10.1186/s40814-023-01416-w] [Medline: 37990337]
27. Bowen DJ, Kreuter M, Spring B, Cofta-Woerpel L, Linnan L, Weiner D, et al. How we design feasibility studies. Am J Prev Med 2009;36(5):452-457 [FREE Full text] [doi: 10.1016/j.amepre.2009.02.002] [Medline: 19362699]
28. Abdelfatah E, Ramos-Santillan V, Cherkassky L, Cianchetti K, Mann G. High risk, high reward: frailty in colorectal cancer surgery is associated with worse postoperative outcomes but equivalent long-term oncologic outcomes. Ann Surg Oncol 2023;30(4):2035-2045. [doi: 10.1245/s10434-022-12970-7] [Medline: 36648616]
29. Low CA, Danko M, Durica KC, Kunta AR, Mulukutla R, Ren Y, et al. A real-time mobile intervention to reduce sedentary behavior before and after cancer surgery: usability and feasibility study. JMIR Perioper Med 2020;3(1):e17292 [FREE Full text] [doi: 10.2196/17292] [Medline: 33393915]
30. Jonker LT, Plas M, de Bock GH, Buskens E, van Leeuwen BL, Lahr MMH. Remote home monitoring of older surgical cancer patients: perspective on study implementation and feasibility. Ann Surg Oncol 2021;28(1):67-78 [FREE Full text] [doi: 10.1245/s10434-020-08705-1] [Medline: 32602060]
31. Haveman ME, van Melzen R, Schuurmann RC, Hermens HJ, Tabak M, de Vries JPPM. Feasibility and patient's experiences of perioperative telemonitoring in major abdominal surgery: an observational pilot study. Expert Rev Med Devices 2022;19(6):515-523 [FREE Full text] [doi: 10.1080/17434440.2022.2108703] [Medline: 35975601]
32. Scheenstra B, Bongers BC, Broeders B, Imkamp M, Van Susante L, Kietselaer B, et al. Reasons and predictors of non-participation in a personalized digital prehabilitation care trial for patients undergoing elective cardiothoracic surgery. Interdiscip Cardiovasc Thorac Surg 2023;37(2):ivad123 [FREE Full text] [doi: 10.1093/icvts/ivad123] [Medline: 37486278]
33. van der Velde M, Valkenet K, Geleijn E, Kruisselbrink M, Marsman M, Janssen LM, et al. Usability and preliminary effectiveness of a preoperative mHealth app for people undergoing major surgery: pilot randomized controlled trial. JMIR Mhealth Uhealth 2021;9(1):e23402 [FREE Full text] [doi: 10.2196/23402] [Medline: 33410758]
34. Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby JM, et al. A new framework for developing and evaluating complex interventions: update of medical research council guidance. BMJ 2021;374:n2061 [FREE Full text] [doi: 10.1136/bmj.n2061] [Medline: 34593508]
35. Leenen JPL, Ardesch V, Patijn G. Remote home monitoring of continuous vital sign measurements by wearables in patients discharged after colorectal surgery: observational feasibility study. JMIR Perioper Med 2023;6:e45113 [FREE Full text] [doi: 10.2196/45113] [Medline: 37145849]
36. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]

Abbreviations

BORG 6-20: Borg Rating of Perceived Exertion
PROMIS: Patient-Reported Outcomes Measurement Information System
QoR-15: Quality of Recovery-15
STROBE: STrengthening the Reporting of OBservational studies in Epidemiology
USE: Usefulness, Satisfaction, and Ease of Use

Edited by N Rohatgi; submitted 11.10.24; peer-reviewed by H Gandhi, L Fortuna; comments to author 18.11.24; revised version received 13.12.24; accepted 15.12.24; published 25.02.25.

Please cite as:
Talen AD, Leenen JPL, van der Sluis G, Oldenhuis HKE, Klaase JM, Patijn GA
Feasibility of a Comprehensive eCoach to Support Patients Undergoing Colorectal Surgery: Longitudinal Observational Study
JMIR Perioper Med 2025;8:e67425
URL: https://periop.jmir.org/2025/1/e67425
doi:10.2196/67425
PMID:

Virtual Reality Exposure Therapy and Patient Education for Preoperative Anxiety in Pediatrics: Randomized Controlled Trial

Sebastian Amaya[1,2,3*], MD; Sidhant Kalsotra[3*], MD; Sibelle Aurelie Yemele Kitio[3*], MPH, MS; Joseph Drew Tobias[3,4*], MD; Brittany Willer[3,4*], MD

[1]Department of Anesthesiology, School of Medicine, Yale University, 333 Cedar St, New Haven, New Haven, CT, United States

[2]Anesthesiology and Critical Care Interest Group, School of Medicine, Universidad El Bosque, Bogotá, Colombia

[3]Department of Anesthesiology and Pain Medicine, Nationwide Children's Hospital, Columbus, OH, United States

[4]Department of Anesthesiology and Pain Medicine, The Ohio State University Wexner Medical Center, Columbus, OH, United States

[*]all authors contributed equally

Corresponding Author:
Sebastian Amaya, MD
Department of Anesthesiology, School of Medicine, Yale University, 333 Cedar St, New Haven, New Haven, CT, United States

Abstract

Background: The perioperative environment is complex and may be challenging for patients and guardians to navigate. The emotional burden and stressors inherent to the perioperative process commonly result in preoperative anxiety. Many studies have demonstrated the usefulness of virtual reality (VR) in various patient populations.

Objective: The aim of this study is to evaluate the impact of a VR-based preoperative education tool on anxiety levels in pediatric patients undergoing ambulatory ear, nose, and throat surgery, as well as in their guardians.

Methods: We performed a single-center prospective randomized controlled trial including children 6 - 12 years of age, presenting for ambulatory tonsillectomy and/or adenoidectomy, with or without bilateral ear tube insertion. The patients were randomized to receive VR instruction of the perioperative workflow or standard preoperative experience (non-VR). The primary outcome was patient and guardian preoperative anxiety, as measured by the 6-item State-Trait Anxiety Inventory.

Results: The study cohort included 107 patient-guardian dyads—51 in the intervention (VR) group and 56 in the control (non-VR) group. Baseline characteristics between the study and control groups were comparable; however, patients in the control group were more likely to report feeling upset compared to the VR group. The VR intervention was associated with reduced preoperative anxiety in patients and guardians compared to the control group. Patients exposed to the VR intervention had higher odds of feeling calm (OR 4.95, 95% CI 2.32 - 10.61; *P*<.001) and lower odds of feeling worried (OR 0.25, 95% CI 0.12 - 0.53; *P*<.001) compared to the control group. Similarly, guardians in the VR group had higher odds of feeling calm (OR 3.55, 95% CI 1.69 - 7.49; *P*=.001) and lower odds of feeling worried (OR 0.45, 95% CI 0.22 - 0.93; *P*=.03) compared to the control group. Both patients and guardians exposed to VR were significantly less likely to have moderate or high levels of preoperative anxiety than the control group (patients: OR 0.15, 95% CI 0.05 - 0.41, *P*<.001; guardians: OR 0.14, 95% CI 0.06 - 0.38, *P*<.001).

Conclusions: VR exposure may be effective in reducing pediatric and guardian anxiety. VR may be a suitable alternative to pharmacologic anxiolysis and future studies should compare the effect to premedication techniques.

Trial Registration: ClinicalTrials.gov NCT05008107; https://clinicaltrials.gov/study/NCT05008107

(JMIR Perioper Med 2025;8:e73392) doi:10.2196/73392

KEYWORDS

anesthesiology; pediatrics; anxiety; virtual reality; exposure therapy

Introduction

The perioperative environment is complex and may be challenging for patients and guardians to navigate. The emotional burden and stressors inherent to the perioperative process commonly result in preoperative anxiety. Up to 75% of children and 50% - 75% of guardians experience preoperative anxiety [1,2]. Indeed, a mutual influence between the patient-guardian dyad can exacerbate or ameliorate the anxiety [3]. Preoperative anxiety is not benign; it is associated with difficult induction, an increased requirement for induction agents, intraoperative hemodynamic instability, emergence delirium, poorer pain control, prolonged recovery, and a greater risk of postoperative complications [4,5]. Additionally, preoperative anxiety may lead to behavioral changes that persist beyond the postoperative period, with up to 88% of children experiencing sleep and eating disturbances, enuresis, temper tantrums, or separation anxiety [6,7]. Though these maladaptive

behaviors are most common in the first 2 weeks following anesthesia, they may persist for up to a year [7-9].

Educating the child and guardian about anesthesia, surgery, and postoperative recovery has been shown to improve preoperative anxiety [10]. Traditionally, preoperative education is provided to the guardian in written form in the surgery clinic, by a nurse phone call in the days leading up to surgery, and on the day of the procedure by preoperative nurses and anesthesiologists. However, production pressures, limited time, conflicting or inaccurate internet-based resources, and poor health literacy impede optimal coaching and effective communication with patients and their guardians in these settings [11-13]. Furthermore, despite best efforts to provide preoperative education, most children experience significant fear of the unfamiliar environment in the operating room.

Virtual reality (VR) presents a unique opportunity to support pediatric patients during the perioperative period by providing an immersive and engaging way to familiarize them with the often-intimidating process of anesthesia induction. Previous studies have demonstrated the effectiveness of various VR technologies in reducing pain and anxiety among children in stressful clinical settings [14-18]. Building on this foundation, our study aimed to evaluate the impact of a VR-based educational tool on preoperative anxiety in both children and their guardians. We hypothesized that exposure to the VR experience would result in reduced anxiety levels for children and guardians prior to surgery.

Methods

Overview

Following institutional review board ethical approval, we performed a single-center prospective randomized controlled trial including children 6 - 12 years of age, presenting for ambulatory tonsillectomy and/or adenoidectomy, with or without bilateral ear tube insertion. We followed CONSORT (Consolidated Standards of Reporting Trials) guidelines for reporting randomized controlled trials and registered the study on ClinicalTrials.gov (NCT05008107). We consented and enrolled patients from October 2021 to September 2023. Patients were excluded if they were scheduled for inpatient admission, had undergone a procedure at our facility within the previous year (to limit the influence of prior familiarity with the perioperative process), were scheduled to undergo additional procedures alongside their primary surgery, or had been prescribed premedication for anxiolysis. VR interventions may pose challenges for individuals with sensory sensitivities, posttraumatic stress disorder, or other neurocognitive conditions due to the immersive and potentially overstimulating nature of the technology. To mitigate these risks, patients with a history of severe motion sickness, visual impairment, or known sensory processing disorders were excluded from participation. Additionally, preprocedural screening included a brief clinical interview to assess conditions such as posttraumatic stress disorder or anxiety disorders that could be exacerbated by immersive experiences. Given that the VR instructional tool was only available in English, we limited our study to children and guardians using English as their language of care.

The patient and their guardian arrived to the hospital approximately 1 hour prior to the scheduled procedure. Following registration, a preoperative nurse confirmed medications and allergies, took vital signs, and brought the patient and guardians to the preoperative holding area. Enrollment was done approximately 30 - 45 minutes before the procedure, and informed consent was obtained from the patient's guardian by a member of the research team. Patients were then block randomized to either receive VR exposure therapy or serve as controls (no intervention). Research team members were not blinded to group allocation. The primary outcome was patient anxiety. Baseline anxiety was assessed using the 6-item State-Trait Anxiety Inventory (6-STAI), a validated tool in pediatric populations [19]. Both patients and their guardians completed the 6-STAI on tablet devices provided by the research team (Table S1 in Multimedia Appendix 1) prior to the intervention. Research staff were not privy to individual responses. Patients assigned to the VR group viewed a proprietary instructional 5-minute video [20] using an Oculus (Meta) headset in the preoperative holding area. Guardians then viewed the same video. Following the VR exposure, the 6-STAI was readministered to both patients and guardians in the VR group to assess postintervention anxiety. The control group did not complete a second anxiety assessment, as only a brief period had elapsed between initial and potential follow-up measurements, making meaningful change unlikely. Following VR exposure (or in the absence of VR for the control group), the anesthesiologist and surgeon met with the patient and guardian to discuss the anesthetic plan and procedure. These discussions were not standardized. An OR nurse retrieved the patient just prior to surgery and separation from the guardian occurred just outside the doors to the operating room.

VR Exposure

The VR program is a 360° video (non–computer generated) of a generic surgical center that provides an immersive first-person experience of the perioperative process. The video is narrated by a child and begins in the waiting room, welcoming the patient for their procedure. The patient can explore the waiting room and reception area as the narrator highlights the nil per os requirements that the patient should have abided by. The patient then enters a consultation room to meet with a pediatric anesthesiologist. The narrator explains the role of the pediatric anesthesiologist. The patient then takes a ride on a gurney through the operating room halls to the operating room. The patient can explore the room as the anesthesiologist introduces various tools used to provide anesthesia, including the face mask. The patient then lays back on the operating room table and the narrator introduces the blood pressure cuff, pulse oximeter, and electrocardiogram leads. The anesthesiologist places the mask over the patient's face to induce anesthesia. Finally, the patient finds themselves in the recovery room waking up with a nurse and anesthesiologist.

Statistical Analysis

We conducted statistical analysis to compare demographic characteristics and outcome measures between the VR and non-VR groups. Descriptive statistics were calculated for demographic characteristics within each group. The

Shapiro-Wilk test was used to assess the assumption of normality for continuous variables. Since continuous variables did not adhere to a normal distribution, they were presented as median and interquartile ranges. Group comparisons for continuous variables were made using the Mann-Whitney *U* test. Categorical variables were summarized as counts and percentages, and appropriate statistical tests, such as the X^2 test or Fisher exact test, were used for group comparisons. An ordinal logistic regression analysis was conducted to compare each of the 6-STAI component measures between the VR and non-VR groups, after testing for the proportionality odds assumption. This involved estimating odds ratios (OR) and 95% CIs to assess the association between the VR exposure and preoperative anxiety in pediatrics. Additionally, based on the validated correlation between the 6-STAI and the full 20-item version—as established by Marteau and Bekker [21]—we calculated a prorated STAI score by multiplying the 6-STAI score by a factor of 20/6. This yielded an estimated full-scale STAI score ranging from 20 to 80 [21]. We then categorized the prorated STAI scores into three groups: low anxiety (20-37), moderate anxiety (38-44), and high anxiety (45-80). Finally, we performed an ordinal logistic regression of the prorated STAI scores to assess the association between the VR exposure and preoperative anxiety. All statistical analyses were performed using Stata (version 18; StataCorp). A *P* value <.05 was adopted to determine statistical significance in all analyses.

A post hoc power analysis was conducted using the prorated STAI scores (20 - 80 range). With 51 participants in the VR group and 56 in the control group, the study had approximately 82% power to detect a clinically meaningful difference of 5 points in anxiety scores. This calculation assumed a standard deviation of 8.9 (based on a previous study by Rizzo et al [22]), a 2-sided α level of .05, and a 2-sample comparison.

Ethical Considerations

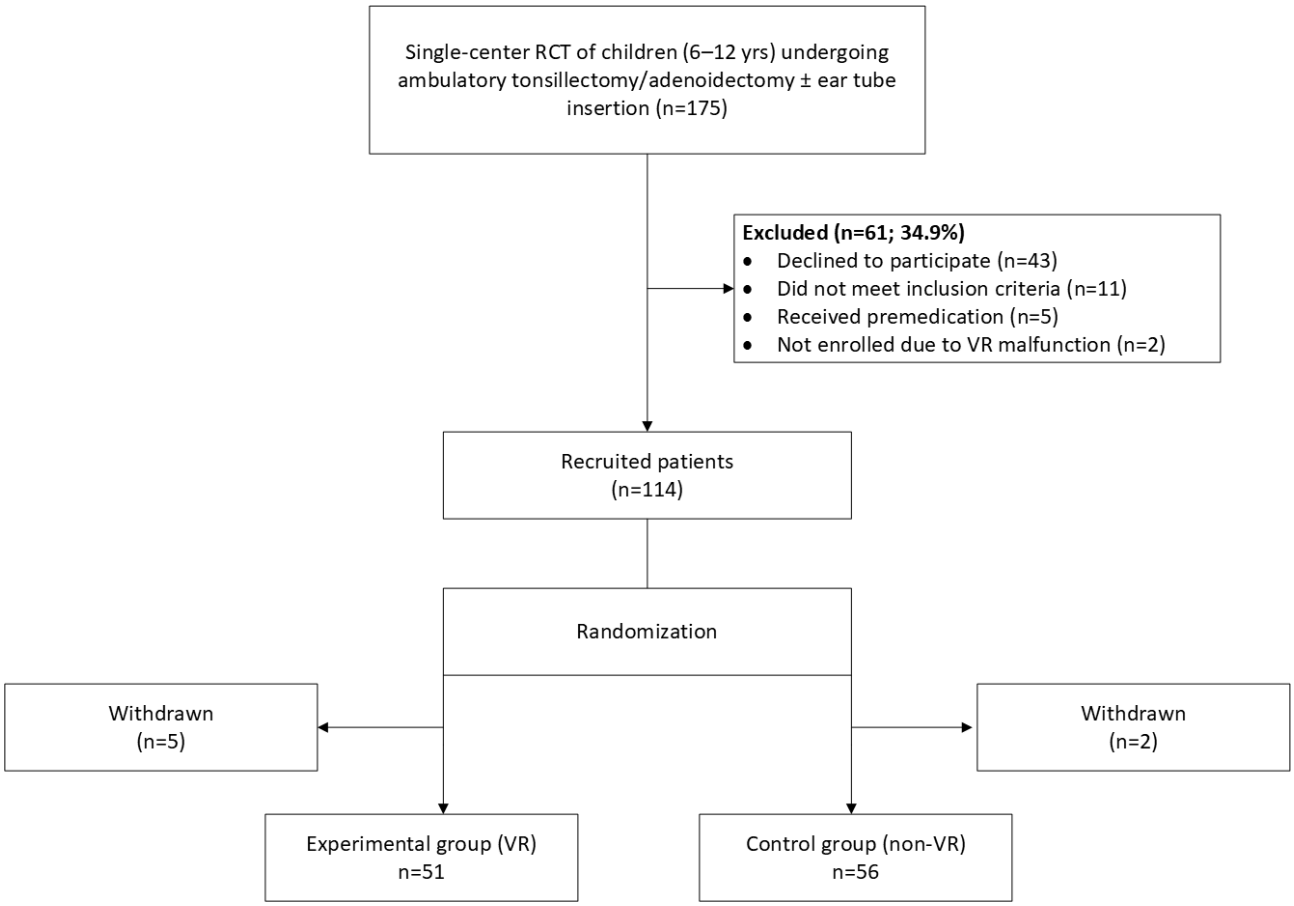
The study was approved by the ethical committee at Nationwide Children's Hospital. Patients' verbal assent and legal guardians' signed consent were obtained. No compensation was provided to participants.

Results

Study Population Characteristics

Of the 175 children assessed for eligibility, 107 patients met the inclusion criteria and consented to participate (Figure 1).

Figure 1. Patient selection flowchart. RCT: randomized controlled trial; VR: virtual reality.



Fifty-one children were randomized to the VR group and 56 to the control group. The average age of children in the study was 8 years, with no significant difference observed between the VR and non-VR groups. The majority of children were female (n=60, 56%) and non-Hispanic White (n=79, 76%). Baseline characteristics between the study and control groups were comparable (Table 1).

Table . Participant demographic data (N=107).

Characteristics	Entire cohort	Non-VR[a]	VR	*P* value[b]
Study population, n (%)	107 (100)	56 (52.3)	51 (47.7)	
Age (years), median (IQR)	8 (7-9)	8 (7-10)	8 (7-9)	.91
Gender, n (%)				.82
Female	60 (56.1)	32 (57.1)	28 (54.9)	
Male	47 (43.9)	24 (42.9)	23 (45.1)	
Race and ethnicity, n (%)				.44
Hispanic	4 (3.7)	3 (5.4)	1 (2)	
Non-Hispanic Black	17 (15.9)	8 (14.3)	9 (17.6)	
Non-Hispanic White	79 (73.8)	43 (76.8)	36 (70.6)	
Other/unknown	7 (6.5)	2 (3.6)	5 (9.8)	
Weight (kg), median (IQR)	31.3 (24.6-42.0)	32.3 (26.5-43.7)	30.9 (23.7-38.2)	.34
Height (cm), median (IQR)	131.1 (125.3-139.3)	131.1 (125.4-141.6)	131.8 (124-139)	.78
BMI (kg/m^2), median (IQR)	18.0 (15.6-22.8)	18.8 (15.5-23.5)	17.2 (15.8-20.8)	.18

[a]VR: virtual reality.

[b]Group comparison was made using the X^2 test or Fisher exact test for categorical variables and Mann-Whitney *U* test for continuous variables.

Baseline (Preintervention) Anxiety

Patients in the VR and non-VR groups had similar preintervention anxiety scores across most STAI components, including feeling calm (*P*=.57), tense (*P*=.28), relaxed (*P*=.22), content (*P*=.27), and worried (*P*=.14). However, children in the non-VR group were significantly more likely to feel upset at baseline than those in the VR group (*P*=.01; Table S1 in Multimedia Appendix 1).

There was no difference between guardian anxiety scores at baseline between the VR and non-VR groups (Table S2 in Multimedia Appendix 1).

Effect of VR on Anxiety by 6-STAI Component

The results of the ordinal logistic regression analysis (Figure 2) revealed that, for patients, the estimated cumulative odds of feeling very calm (OR 4.95, 95% CI 2.32 - 10.61; *P*<.001), relaxed (OR 4.23, 95% CI 2.00 - 9.05; *P*<.001), and content (OR 2.13, 95% CI 1.04 - 4.34; *P*=.04) were higher for those who used VR compared to those who did not. Conversely, the estimated cumulative odds of feeling tense (OR 0.27, 95% CI 0.12 - 0.61; *P*=.001), upset (OR 0.14, 95% CI 0.04 - 0.50; *P*=.003), and worried (OR 0.25, 95% CI 0.12 - 0.53; *P*<.001) were significantly lower in patients who used VR. Similarly, for guardians, those who received VR reported higher odds of feeling very calm (OR 3.55, 95% CI 1.69 - 7.49; *P*=.001), relaxed (OR 4.02, 95% CI 1.90 - 8.50; *P*<.001), and content (OR 3.34, 95% CI 1.56 - 7.16; *P*=.002) compared to those who did not receive VR. Additionally, they had lower odds of feeling tense (OR 0.42, 95% CI 0.19 - 0.91; *P*=.03) and worried (OR 0.45, 95% CI 0.22 - 0.93; *P*=.03). However, no statistically significant difference was found for feeling less upset (OR 0.25, 95% CI 0.05 - 1.22; *P*=.09) between guardians who received VR and those who did not (Figure 3).

Figure 2. Ordinal logistic regression analysis examining the impact of virtual reality on patient preoperative anxiety.

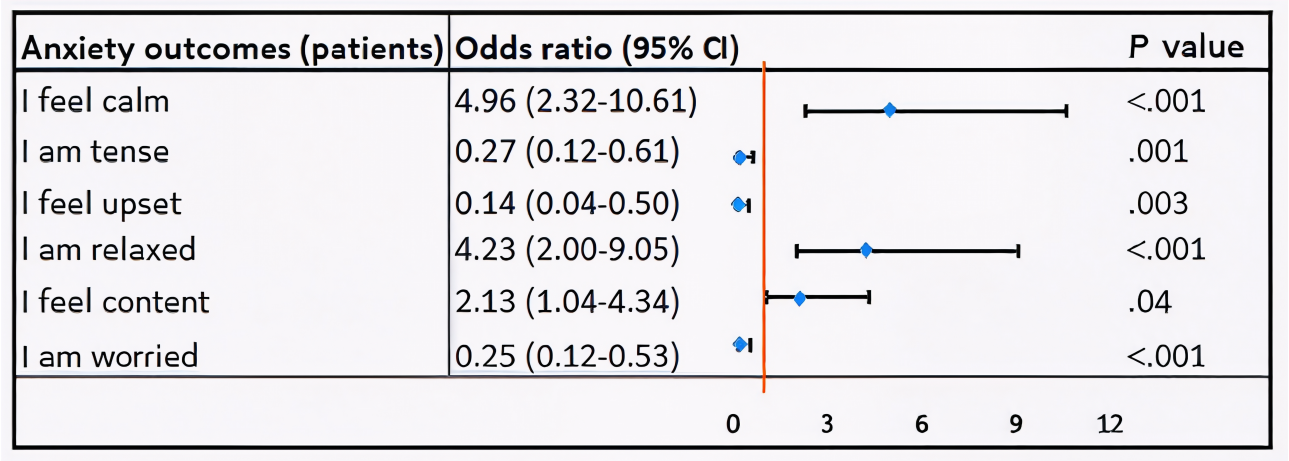
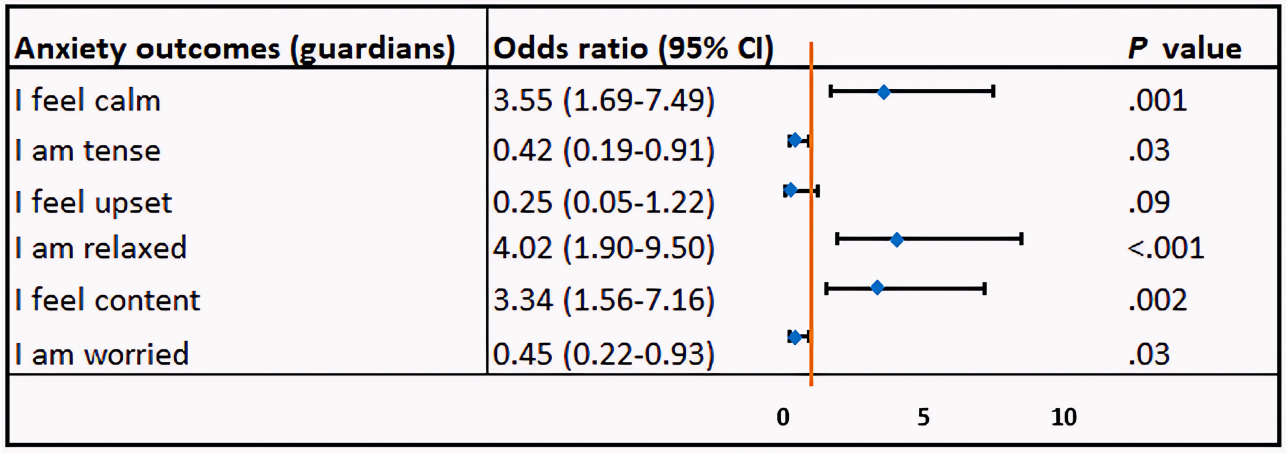


Figure 3. Ordinal logistic regression analysis examining the impact of virtual reality on guardian preoperative anxiety.



Effect of VR on Prorated STAI Anxiety Level

The median prorated STAI scores were higher in the control group for both patients (37) and their guardians (40) compared those who received VR (patients: 33; guardians: 33). Among the VR group, median anxiety scores decreased following exposure: for patients, from 33 (pre-VR) to 30 (post-VR), and for guardians, from 33 to 27. VR exposure was also associated with a reduced proportion of patients with high anxiety from 25.5% to 9.8% and moderate anxiety from 7.8% to 2%. Similarly, the proportion of guardians reporting high anxiety dropped from 32.7% to 6.4%, and moderate anxiety dropped from 12.2% to 8.5% (Tables S3 and S4 in Multimedia Appendix 1).

Effect of VR on the Odds of a High or Moderate Anxiety Level

The additional analysis of the prorated STAI score showed that after the intervention, patients exposed to VR had 85% lower odds of having high or moderate preoperative anxiety than those in the control group (OR 0.15, 95% CI 0.05 - 0.41; $P<.001$). A similar pattern was observed in guardians, where those exposed to VR had significantly lower odds of high or moderate anxiety relative to those not exposed to VR (OR 0.14, 95% CI 0.06 - 0.38; $P<.001$).

Discussion

In this prospective randomized controlled trial, we demonstrated the ability of VR-based preoperative instruction to decrease anxiety of children undergoing ambulatory tonsillectomy. Specifically, the patients and guardians exposed to the VR tool were significantly less likely to experience moderate or high preoperative anxiety than those in the control group. They were also more likely to report feeling calm, relaxed, and content prior to surgery than those who were not exposed to VR and were less likely to report feeling tense, upset, and worried. These findings may represent an advance in pediatric perioperative care—preoperative anxiety remains a hurdle for pediatric anesthesiologists and their patients, with the optimal panacea yet to be identified. Many anesthesiologists rely on pharmacologic interventions to ease the child's stress. VR technology may offer an effective alternative to pharmacologic premedication, without the risk of adverse effects.

Much of the existing pediatric VR research centers on distraction-based techniques, which aim to divert attention away from stressful stimuli using immersive and engaging VR experiences [23-26]. This approach has shown effectiveness in various clinical settings, including peripheral intravenous placement, anesthesia induction, and preoperative preparation [17,27,28]. In contrast, our study used VR-based exposure therapy, which uses virtual environments to simulate a patient's specific stressor—in this case, the perioperative setting—to promote familiarity and reduce anxiety [29]. Research on VR exposure therapy for preoperative anxiety remains limited. For example, Eijlers et al [30] found no significant reduction in preoperative anxiety for children or guardians using VR exposure therapy compared to those who did not use VR; however, children in the VR group required fewer rescue analgesics postoperatively compared to controls. Conversely, Ryu et al [31] reported significantly lower preoperative anxiety and improved anesthesia induction compliance in children who received VR exposure therapy compared to those in the control group.

Anxiety reduction following our VR exposure intervention can likely be attributed to both desensitization—reducing anticipatory fear and distress—and information delivery [32]. Studies have shown that educational information improves both anxiety and guardian satisfaction [30,33,34]. Furthermore, data suggest that children desire more information about the perioperative experience, including details of the surgical environment, presence of their guardian, conduct of anesthesia, and expectations for postoperative pain. VR exposure therapy can provide the desired information in a fun and interactive platform [35].

Implementing VR in the perioperative environment is both feasible and potentially cost-effective, particularly when thoughtfully integrated into existing clinical workflows. VR interventions can be administered by either nurses or child life specialists, depending on institutional resources. Both roles are well-suited to deliver the experience with minimal additional training. The intervention itself is brief—typically requiring only about 5 minutes—making it easy to incorporate into standard preoperative routines. From an infrastructure

perspective, the requirements are modest—ideally, there should be one VR headset per operating room to ensure consistent access for all patients. Most commercial VR headsets cost under US $500, making them a realistic investment for hospitals. Maintenance is straightforward, with simple hygiene protocols such as disinfecting the headset between uses with alcohol swabs. Moreover, the flexibility of the intervention allows it to be implemented not only on the day of the surgery but also in preoperative clinics or during earlier surgical consultations. This expanded access provides families with multiple opportunities to familiarize themselves with the perioperative process, supporting anxiety reduction while maintaining clinical efficiency.

It is important to note the limitations of our study. First, although an a priori sample size calculation was not performed, a post hoc power analysis confirmed that the study was adequately powered to detect a clinically meaningful difference in preoperative anxiety between groups. Our study was limited to school-aged children undergoing tonsillectomy and/or adenoidectomy. However, preoperative anxiety afflicts children of all ages and varies with type of surgery, anesthetic, and expected disposition [4]. Therefore, our findings may not be generalizable to all populations of children undergoing surgery. Additionally, the STAI questionnaire, which is the most frequently used anxiety assessment, is quite extensive and may lead to response bias [4]. Postintervention anxiety was only assessed in the VR group, introducing potential confounding. Anxiety is dynamic; therefore, it is conceivable that STAI scores could fluctuate after elapsed time or following additional interactions with health care providers prior to the surgery. Although virtual reality exposure therapy appears to be an effective tool for reducing pediatric preoperative anxiety, it has yet to be directly compared to the current gold standard for anxiolysis: midazolam. Future studies should assess the comparative effectiveness of pharmacologic interventions versus VR-based approaches in reducing preoperative anxiety. Additionally, the potential impact of VR exposure on postoperative pain and analgesic requirements should be explored. To enhance accessibility and equity, VR tools should be developed for patients who receive care in languages other than English. With these considerations in mind, the findings of this study suggest that VR exposure therapy may be a valuable intervention for reducing anxiety in both pediatric patients and their guardians.

In conclusion, this single-center randomized controlled trial demonstrated that VR exposure therapy may have a beneficial effect on reducing preoperative anxiety in both pediatric patients and their guardians. By providing an immersive, nonpharmacologic, and engaging preparatory experience, VR may serve as a valuable adjunct to traditional perioperative care. These findings suggest that VR has the potential to enhance the surgical experience by addressing emotional well-being in a child- and family-centered manner. Although further research is needed, this study contributes to a growing body of evidence supporting the role of digital therapeutics in pediatric perioperative care.

Acknowledgments

Dr Vidya Raman and NDS Kalstars created the virtual reality instructional video used in this study. Internal funding was provided by the authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplemental tables showing preintervention and postintervention anxiety in patients and guardians as assessed by the 6-State-Trait Anxiety Inventory
[DOCX File, 24 KB - periop_v8i1e73392_app1.docx]

Checklist 1
CONSORT checklist.
[PDF File, 175 KB - periop_v8i1e73392_app2.pdf]

References

1. Getahun AB, Endalew NS, Mersha AT, Admass BA. Magnitude and factors associated with preoperative anxiety among pediatric patients: cross-sectional study. Pediatric Health Med Ther 2020;11:485-494. [doi: 10.2147/PHMT.S288077] [Medline: 33364873]
2. Thompson N, Irwin MG, Gunawardene WM, Chan L. Pre-operative parental anxiety. Anaesthesia 1996 Nov;51(11):1008-1012. [doi: 10.1111/j.1365-2044.1996.tb14992.x] [Medline: 8943589]
3. Cui X, Zhu B, Zhao J, Huang Y, Luo A, Wei J. Parental state anxiety correlates with preoperative anxiety in Chinese preschool children. J Paediatr Child Health 2016 Jun;52(6):649-655. [doi: 10.1111/jpc.13176] [Medline: 27144949]
4. Friedrich S, Reis S, Meybohm P, Kranke P. Preoperative anxiety. Curr Opin Anaesthesiol 2022 Dec 1;35(6):674-678. [doi: 10.1097/ACO.0000000000001186] [Medline: 36131642]

5. Clausen NG, Madsen D, Rosenkilde C, Hasfeldt-Hansen D, Larsen LG, Hansen TG. The use of tablet computers to reduce preoperative anxiety in children before anesthesia: a randomized controlled study. J Perianesth Nurs 2021 Jun;36(3):275-278. [doi: 10.1016/j.jopan.2020.09.012] [Medline: 33637409]
6. Sullivan V, Sullivan DH, Weatherspoon D. Parental and child anxiety perioperatively: relationship, repercussions, and recommendations. J Perianesth Nurs 2021 Jun;36(3):305-309. [doi: 10.1016/j.jopan.2020.08.015] [Medline: 33653615]
7. Kain ZN, Mayes LC, O'Connor TZ, Cicchetti DV. Preoperative anxiety in children. Arch Pediatr Adolesc Med 1996 Dec 1;150(12):1238. [doi: 10.1001/archpedi.1996.02170370016002]
8. Jain S, Patel S, Arora KK, Sharma A. A comparative study on effectiveness of parental presence versus sedative premedication for reducing anxiety in children undergoing general anesthesia. Int J Appl Basic Med Res 2023;13(2):101-105. [doi: 10.4103/ijabmr.ijabmr_636_22] [Medline: 37614833]
9. Yao J, Gong H, Zhao X, Peng Q, Zhao H, Yu S. Parental presence and intranasal dexmedetomidine for the prevention of anxiety during anesthesia induction in children undergoing tonsillectomy and/or adenoidectomy surgery: a randomized controlled trial. Front Pharmacol 2022;13(1015357):1015357. [doi: 10.3389/fphar.2022.1015357] [Medline: 36601054]
10. Kain ZN, Caldwell-Andrews AA, Mayes LC, et al. Family-centered preparation for surgery improves perioperative outcomes in children: a randomized controlled trial. Anesthesiology 2007 Jan;106(1):65-74. [doi: 10.1097/00000542-200701000-00013] [Medline: 17197846]
11. Nordin AB, Shah SR, Kenney BD. Ambulatory pediatric surgery. Semin Pediatr Surg 2018 Apr;27(2):75-78. [doi: 10.1053/j.sempedsurg.2018.02.003] [Medline: 29548355]
12. Hu X, Pennington BRT, Avidan MS, Kheterpal S, NG, Politi MC. Description of the content and quality of publicly available information on the internet about inhaled volatile anesthesia and total intravenous anesthesia: descriptive study. JMIR Perioper Med 2023 Nov 2;6:e47714. [doi: 10.2196/47714] [Medline: 37917148]
13. Lee A, Chui PT, Gin T. Educating patients about anesthesia: a systematic review of randomized controlled trials of media-based interventions. Anesth Analg 2003 May;96(5):1424-1431. [doi: 10.1213/01.ANE.0000055806.93400.93] [Medline: 12707146]
14. Wong CL, Choi KC. Effects of an immersive virtual reality intervention on pain and anxiety among pediatric patients undergoing venipuncture: a randomized clinical trial. JAMA Netw Open 2023 Feb 1;6(2):e230001. [doi: 10.1001/jamanetworkopen.2023.0001] [Medline: 36795410]
15. Tas FQ, van Eijk CAM, Staals LM, Legerstee JS, Dierckx B. Virtual reality in pediatrics, effects on pain and anxiety: a systematic review and meta-analysis update. Paediatr Anaesth 2022 Dec;32(12):1292-1304. [doi: 10.1111/pan.14546] [Medline: 35993398]
16. Olbrecht VA, O'Conor KT, Williams SE, et al. Transient reductions in postoperative pain and anxiety with the use of virtual reality in children. Pain Med 2021 Nov 26;22(11):2426-2435. [doi: 10.1093/pm/pnab209] [Medline: 34175959]
17. Buyuk ET, Odabasoglu E, Uzsen H, Koyun M. The effect of virtual reality on children's anxiety, fear, and pain levels before circumcision. J Pediatr Urol 2021 Aug;17(4):567. [doi: 10.1016/j.jpurol.2021.04.008] [Medline: 34006462]
18. Olbrecht VA, Williams SE, O'Conor KT, et al. Guided relaxation-based virtual reality versus distraction-based virtual reality or passive control for postoperative pain management in children and adolescents undergoing Nuss repair of pectus excavatum: protocol for a prospective, randomised, controlled trial (FOREVR Peds trial). BMJ Open 2020 Dec 30;10(12):e040295. [doi: 10.1136/bmjopen-2020-040295] [Medline: 33380482]
19. Driscoll KA, Melin J, Lynch KF, Smith LB, Johnson SB. SAI-CH-6: development of a short form of the State Anxiety Inventory for children at-risk for type 1 diabetes. J Pediatr Psychol 2023 Oct 20;48(10):861-869. [doi: 10.1093/jpepsy/jsad057] [Medline: 37698990]
20. Raman VT. NDS Kalstars Pediatric Onboarding Experience. Vimeo. URL: https://vimeo.com/441915457?ref=em-share [accessed 2025-10-24]
21. Marteau TM, Bekker H. The development of a six-item short-form of the state scale of the Spielberger State-Trait Anxiety Inventory (STAI). Br J Clin Psychol 1992 Sep;31(3):301-306. [doi: 10.1111/j.2044-8260.1992.tb00997.x] [Medline: 1393159]
22. Rizzo MG, Costello JP, Luxenburg D, Cohen JL, Alberti N, Kaplan LD. Augmented reality for perioperative anxiety in patients undergoing surgery: a randomized clinical trial. JAMA Netw Open 2023 Aug 1;6(8):e2329310. [doi: 10.1001/jamanetworkopen.2023.29310] [Medline: 37589975]
23. Eijlers R, Utens E, Staals LM, et al. Systematic review and meta-analysis of virtual reality in pediatrics: effects on pain and anxiety. Anesth Analg 2019 Nov;129(5):1344-1353. [doi: 10.1213/ANE.0000000000004165] [Medline: 31136330]
24. Tashjian VC, Mosadeghi S, Howard AR, et al. Virtual reality for management of pain in hospitalized patients: results of a controlled trial. JMIR Ment Health 2017 Mar 29;4(1):e9. [doi: 10.2196/mental.7387] [Medline: 28356241]
25. Eccleston C. Chronic pain as embodied defence: implications for current and future psychological treatments. Pain 2018 Sep;159 Suppl 1(PMID):S17-S23. [doi: 10.1097/j.pain.0000000000001286] [Medline: 30113943]
26. Spiegel B, Fuller G, Lopez M, et al. Virtual reality for management of pain in hospitalized patients: a randomized comparative effectiveness trial. PLoS One 2019;14(8):e0219115. [doi: 10.1371/journal.pone.0219115] [Medline: 31412029]

27. Gold JI, SooHoo M, Laikin AM, Lane AS, Klein MJ. Effect of an immersive virtual reality intervention on pain and anxiety associated with peripheral intravenous catheter placement in the pediatric setting: a randomized clinical trial. JAMA Netw Open 2021 Aug 2;4(8):e2122569. [doi: 10.1001/jamanetworkopen.2021.22569] [Medline: 34432011]
28. Jung MJ, Libaw JS, Ma K, Whitlock EL, Feiner JR, Sinskey JL. Pediatric distraction on induction of anesthesia with virtual reality and perioperative anxiolysis: a randomized controlled trial. Anesthesia & Analgesia 2021;132(3):798-806. [doi: 10.1213/ANE.0000000000005004]
29. Krzystanek M, Surma S, Stokrocka M, et al. Tips for effective implementation of virtual reality exposure therapy in phobias-a systematic review. Front Psychiatry 2021;12:737351. [doi: 10.3389/fpsyt.2021.737351] [Medline: 34621197]
30. Eijlers R, Dierckx B, Staals LM, et al. Virtual reality exposure before elective day care surgery to reduce anxiety and pain in children: a randomised controlled trial. Eur J Anaesthesiol 2019 Oct;36(10):728-737. [doi: 10.1097/EJA.0000000000001059] [Medline: 31356373]
31. Ryu JH, Park SJ, Park JW, et al. Randomized clinical trial of immersive virtual reality tour of the operating theatre in children before anaesthesia. Br J Surg 2017 Nov;104(12):1628-1633. [doi: 10.1002/bjs.10684] [Medline: 28975600]
32. Chiu PL, Li H, Yap KYL, Lam KMC, Yip PLR, Wong CL. Virtual reality-based intervention to reduce preoperative anxiety in adults undergoing elective surgery: a randomized clinical trial. JAMA Netw Open 2023 Oct 2;6(10):e2340588. [doi: 10.1001/jamanetworkopen.2023.40588] [Medline: 37906193]
33. Sjöling M, Nordahl G, Olofsson N, Asplund K. The impact of preoperative information on state anxiety, postoperative pain and satisfaction with pain management. Patient Educ Couns 2003 Oct;51(2):169-176. [doi: 10.1016/s0738-3991(02)00191-x] [Medline: 14572947]
34. Jlala HA, French JL, Foxall GL, Hardman JG, Bedforth NM. Effect of preoperative multimedia information on perioperative anxiety in patients undergoing procedures under regional anaesthesia. Br J Anaesth 2010 Mar;104(3):369-374. [doi: 10.1093/bja/aeq002] [Medline: 20124283]
35. Mendoza BA, Fortier MA, Trinh LN, Schmid LN, Kain ZN. Factors impacting parental and child satisfaction in the perioperative setting. Paediatr Anaesth 2021 Sep;31(9):932-943. [doi: 10.1111/pan.14236] [Medline: 34096658]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
OR: odds ratio
STAI-6: 6-item State-Trait Anxiety Inventory
VR: virtual reality

Edited by N Rohatgi; submitted 03.03.25; peer-reviewed by L Bacalhau, T Kim; revised version received 28.09.25; accepted 15.10.25; published 27.10.25.

Please cite as:
Amaya S, Kalsotra S, Yemele Kitio SA, Tobias JD, Willer B
Virtual Reality Exposure Therapy and Patient Education for Preoperative Anxiety in Pediatrics: Randomized Controlled Trial
JMIR Perioper Med 2025;8:e73392
URL: https://periop.jmir.org/2025/1/e73392
doi:10.2196/73392

Use of Virtual Reality in the Pediatric Perioperative Setting and for Induction of Anesthesia: Mixed Methods Pilot Feasibility Study

Yu Tong Huang[1,2], MD; Sofia Addab[1,2], MSc; Gianluca Bertolizio[3], MD; Reggie Hamdy[2,4], MD, MSc; Kelly Thorstad[2], NP, MSc; Argerie Tsimicalis[1,2,5], RN, PhD

[1]Faculty of Medicine and Health Sciences, McGill University, Montreal, QC, Canada

[2]Clinical Research, Shriners Hospitals for Children–Canada, 1003 Decarie Blvd, Montreal, QC, Canada

[3]Division of Pediatric Anesthesia, The Montreal Children's Hospital, McGill University Health Center, Montreal, QC, Canada

[4]Division of Pediatric Surgery, The Montreal Children's Hospital, McGill University Health Center, Montreal, QC, Canada

[5]Faculty of Dental Medicine and Oral Health Sciences, McGill University, Montreal, Canada

Corresponding Author:
Argerie Tsimicalis, RN, PhD
Faculty of Medicine and Health Sciences, McGill University, Montreal, QC, Canada

Abstract

Background: Children commonly experience high levels of anxiety prior to surgery. This distress is associated with postoperative maladaptive behaviors. Virtual reality (VR) is an innovative tool for reducing anxiety and pain during various medical procedures. Previous randomized controlled trials have demonstrated its efficacy in reducing children's anxiety in the preoperative waiting room or during induction.

Objective: The primary aim of this study was to examine the feasibility of VR distraction throughout the perioperative period, from the waiting room until the induction of general anesthesia (GA). Secondary aims were to assess its clinical utility, tolerability, and initial clinical efficacy.

Methods: A mixed methods, concurrent triangulation feasibility trial was piloted at the Shriners Hospitals for Children–Canada. Participants played an interactive VR game throughout the perioperative period, starting from the waiting room until induction. Feasibility was examined with the duration of the VR intervention, recording the number of interruptions, and taking field notes. Clinical utility was assessed using a perception questionnaire. Tolerability was evaluated by the Child Simulator Sickness Questionnaire (CSSQ). Initial clinical efficacy was assessed by the Faces Pain Scale–Revised, Faces Anxiety Scale, Graphic Rating Scale for multidimensional pain, the Induction Compliance Checklist, and the Pediatric Anesthesia Emergence Delirium scale. Quantitative data were supported with field notes and semistructured interviews with patients and parents. Quantitative and qualitative themes were compared via the triangulation protocol to produce final themes.

Results: A total of 39 patients, with a mean age of 11.9 (SD 2.8) years, undergoing elective surgery under GA participated in the study. Stakeholders, including patients, parents, and health care providers, were receptive and willing to adapt to VR. Of the 39 patients, 19 (49%) continued to use VR during transportation and 6 (15%) were induced with VR. Barriers to feasibility included (1) interruptions to VR in 92% (36/39) of patients by health care professionals, (2) unpredictable surgery delays prolonging the duration of the VR intervention (mean 23.1, SD 24.4 minutes; range 5 - 150 minutes), and (3) discontinuation of VR before induction due to mask seal (n=3) and discomfort with supine positioning (n=2). Patients were generally satisfied with VR, deemed it acceptable and easy to use, and would recommend it to others. VR was tolerable with no self-reported simulator sickness (CSSQ: mean 0.01, SD 0.1). The mean Faces Anxiety Score was 1.5 (SD 1.1) at baseline and 0.7 (SD 0.9) during VR.

Conclusions: While VR demonstrated good clinical utility and was well tolerated in the broad perioperative setting, this study highlighted important feasibility barriers in the waiting room and especially during induction of anesthesia, both at the organizational and technical levels. This study highlights several considerations that should be carefully addressed for the successful implementation of perioperative VR.

(JMIR Perioper Med 2025;8:e58905) doi:10.2196/58905

KEYWORDS

virtual reality; augmented reality; mixed reality; extended reality; computer-generated simulation; digital world; virtual environment; anxious; pediatrics; anesthetics; preoperative; feasibility; artificial intelligence; digital health technology; surgery; child care

Introduction

Virtual reality (VR) is an innovative tool for managing anxiety and pain during medical procedures, such as needle insertion, dressing changes, and dental care [1]. Up to 60% of children experience high levels of anxiety at induction of general anesthesia (GA) [2,3]. This distress is associated with a greater risk of postoperative emergence delirium [4,5], disturbed sleep, and behavioral and emotional disturbances [6]. Previous studies and a meta-analysis have demonstrated the efficacy of preoperative operating room (OR) tours by VR in reducing anxiety [7-18]. With regard to VR use during induction of anesthesia, 2 randomized controlled trials (RCTs) by Jung et al [19] and Samnakay et al [20] have demonstrated the efficacy of a VR game for distraction and anxiety reduction compared to standard care or noninferiority to the use of a 2D tablet. The perioperative period is a continuum of multiple moments that can cause anxiety, including awaiting surgery in the waiting room, being transported to the OR, and undergoing induction of anesthesia in the OR. Understanding the VR feasibility for distraction across these different moments would be fundamental to determining if and where VR can be integrated into the overall perioperative patient flow. Integrating VR into the induction of anesthesia may be more technically complex than in other studied settings, such as intravenous (IV) insertions. Hence, the primary objective of this study was to determine the feasibility of using VR for distraction in the perioperative setting, from the preoperative waiting room to induction. Secondary aims were to assess the clinical utility, tolerability, and initial effectiveness of VR in the same time frame.

Methods

Study Design and Setting

A mixed method [21], concurrent triangulation feasibility study was piloted on the OR floor of the Shriners Hospitals for Children–Canada, a university-affiliated, not-for-profit, bilingual, pediatric orthopedic hospital located in Montreal, Quebec, Canada. The OR floor consisted of a preoperative waiting room, 4 ORs, and a postanesthesia care unit (PACU). In this study, 14 anesthesiologists, 17 surgeons, 14 nurses, 3 respiratory therapists, and 4 orderlies were involved in VR use. The study commenced during the COVID-19 pandemic, resulting in parents no longer being allowed in the OR during the anesthesia induction, with recruitment starting in May 2021 and ending in June 2022.

VR-CORE Outcomes for VR2 Trials

The VR Clinical Outcomes Research Experts (VR-CORE) methodological framework guided this VR2 trial, of which the aim is to produce an "initial assessment in the target patient population within a representative clinical setting" [22]. The primary outcome was the feasibility of perioperative VR, consisting of barriers and facilitators to this intervention [22]. Secondary outcomes consisted of: (1) clinical utility, defined as acceptability, ease of use and understanding, satisfaction, and recommendation of the VR intervention; (2) tolerability, which entailed the absence of physical or emotional adverse effects related to VR; and (3) initial clinical efficacy, defined as patients' outcomes of anxiety, pain, and compliance at induction.

Participants

Convenience sampling was used to prospectively recruit participants in the preoperative evaluation clinic. Patients were eligible if they (1) were aged between 5 and 21 years, (2) had a scheduled elective surgery under GA, and (3) could understand French or English. Patients were excluded if they (1) had a cognitive, auditory, or visual impairment preventing VR use or (2) had a history of seizures or epilepsy. Parents or legal guardians were eligible if they were present with the child and were willing to share their perspectives. A sample size of approximately 40 patients was based on a prior feasibility study including at the study site [23] and a similar setting [24]. The sample size aligned with the VR-CORE recommendations [22].

VR Intervention

Participants played a pretested [23,25,26], interactive game with sound, DREAM (Paperplane Therapeutics, Inc) [27], via the Pico Neo 3 headset. DREAM entails patients throwing red balls at balloons, diamonds, and trolls in a fantastical landscape. The game was developed with input from medical professionals and tested at pediatric sites, including the study site [23,25,26]. DREAM was designed for health care use, allowing for (1) reduced speed to prevent VR sickness, (2) one hand for play, (3) head movement to orient the character, (4) aesthetic appeal, and (5) a no "loss" state. One headset was available for this study. The game was not mirrored onto a tablet for parents or clinicians to view to avoid internet connection-related interruptions to gameplay.

Study Procedure

Nursing staff in the preoperative clinic helped identify eligible participants. A member of the research team explained the study to parents and patients and, if agreeable, obtained informed consent and assent during their preoperative appointment, days or weeks before their surgery. On the day of surgery, game instructions were provided, and the headset was fitted to the patient by a researcher in the preoperative waiting room. The VR intervention was offered within the workflow of the study site (Figure 1). At least 5 minutes of gameplay for VR immersion was encouraged before transfer to the OR; however, gameplay was allowed to extend for longer durations if there were OR delays. Patients were encouraged to pause after 30 minutes of screen time to avoid VR sickness. The researcher remained on standby and troubleshooted VR issues. Patients could pause or discontinue VR at any point. Nurses, orderlies, surgeons, and anesthesiologists usually visited patients in the waiting room to perform preoperative tasks and assessments before transfer to the OR. At the study site, patients routinely received Tylenol and did not routinely receive anxiolytics unless deemed necessary. EMLA cream was routinely applied 30 minutes prior to awake IV insertions. Patients were verbally notified when it was time for their surgery and were asked if they wanted to continue VR for transfer to the OR and induction. Patients were preoxygenated with an age-appropriate mask. The method of induction, usually inhalational with sevoflurane at

the study site, was left to the anesthesiologist's discretion. After induction, the headset was removed.

Figure 1. Virtual reality intervention at various perioperative time points: in the waiting room (top), during transport (middle), and during induction (bottom).

Data Collection

After informed consent, baseline sample characteristics were collected from patients and hospital charts. Patient-reported anxiety, pain, and VR sickness were subsequently collected in the waiting room using the Faces Anxiety Scale (FAS) [28], the Faces Pain Scale–Revised (FPS-R) [29], and the Child Simulator Sickness Questionnaire (CSSQ) [30], respectively. During the VR intervention, field notes were taken by a researcher [23]. In the OR, induction compliance was assessed by a researcher via the Induction Compliance Checklist (ICC) [31]. In the PACU, as part of standard practice, the nurses recorded the emergence delirium using the Pediatric Anesthesia Emergence Delirium (PAED) scale [4]; the scores were retrieved

from the chart. Following the surgery, patients retrospectively reported anxiety, pain, and VR sickness experienced during VR use using the FAS, FPS-R, and CSSQ, as well as the Graphic Rating Scale (GRS) [32]. Patient perception was assessed using a modified version of the Patient Perception Questionnaire [33]. These data were collected either immediately after surgery in the PACU, a few days later in the inpatient unit, or weeks to months later at the follow-up appointment. Finally, if agreeable, the patient and their parent (if present) participated in an audio-recorded, semistructured interview, using a previously used interview guide created by the study team [23].

Data Analysis

Descriptive statistics were used to summarize the sociodemographic and instrument data using Microsoft Excel (2016) measures of central tendency and variance, generating a list of key findings set aside for triangulation. Qualitative data analysis was conducted separately through directed content analysis [34] of the field notes and interviews. The themes identified were supported by quotes, observations, and field notes. Triangulation analysis of qualitative and quantitative data led to the identification of meta-themes, which resulted from qualitative and quantitative sources, and themes drawn from one data source [23,35]. Through this process, qualitative and quantitative data were compared and contrasted, resulting in “agreement,” “disagreement,” “silence,” or “complementarity” between the two data source.

Ethical Considerations

This study was performed in accordance with the principles of the Declaration of Helsinki. Approval was granted by the McGill Institutional Review Board (A06-M31-19B). Informed consent was obtained from all participants and legal guardians included in the study. All data were deidentified prior to analysis to maintain participant privacy. Participants received no monetary compensation.

Results

Sample Characteristics

In total, 61 eligible patients were approached for study participation, of which 49 consented or assented to participate. Two patients withdrew on the morning of their scheduled surgery before the VR intervention, as they were no longer interested in using VR. Eight patients were lost to follow-up as their surgery was canceled, and their rescheduled date conflicted with other participants (Figure 2). Overall, 39 patients, with a mean age of 11.9 (SD 2.8) years and a median age of 12 (IQR 10-13.5) years, used VR in the perioperative setting, for a participation rate of 64% (39/61; Table 1). A total of 11 participants and 9 parents agreed to participate in the interview following their surgery. The remaining patients were lost to follow-up after their surgery.

Figure 2. CONSORT (Consolidated Standards of Reporting Trials) flow diagram. OR: operating room; VR: virtual reality.

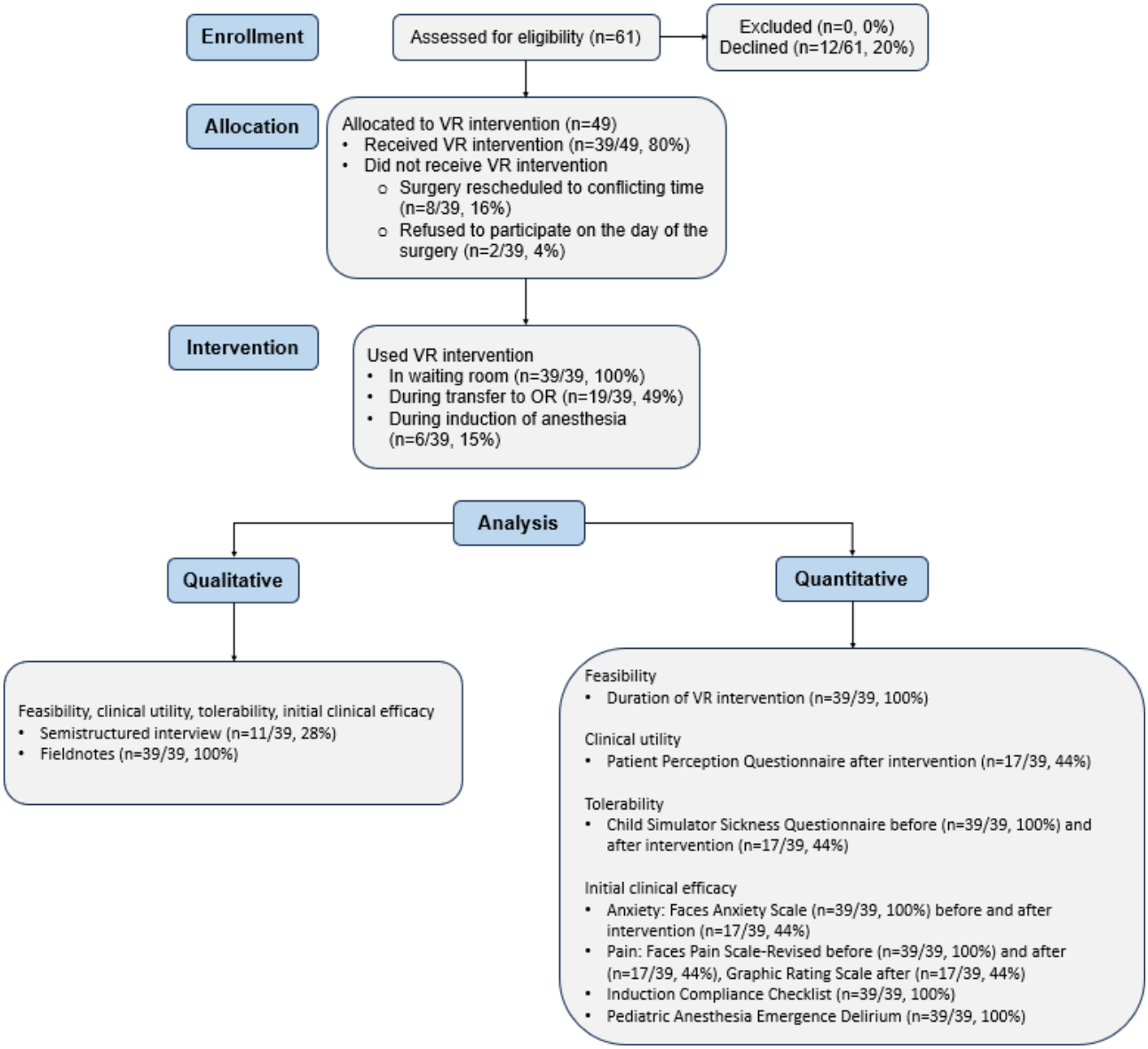


Table . Patient demographics (n=39).

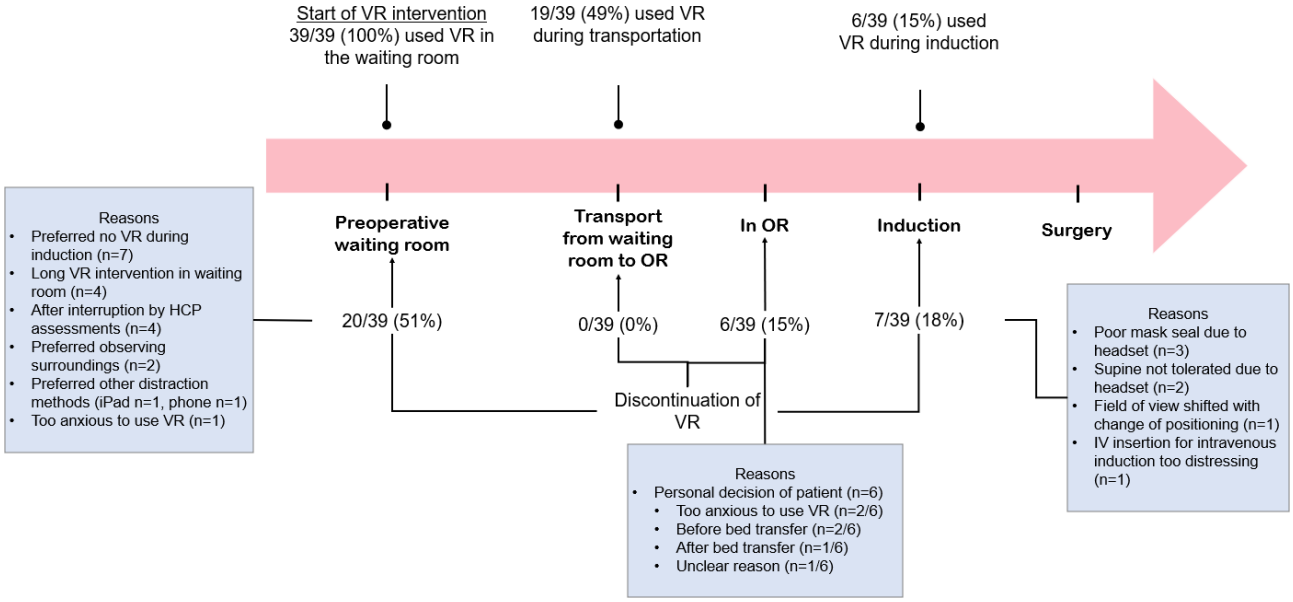
Characteristics	Values
Age (years)	
Mean (SD)	11.9 (2.8)
Median (IQR)	12 (10 - 13.5)
Sex, n (%)	
Male	18 (46)
Female	21 (54)
Race, n (%)	
White	30 (77)
Black	5 (13)
Hispanic	1 (3)
Other	3 (8)
Patients receiving preoperative medications, n (%)	
Tylenol	39 (100)
Midazolam	2 (5)
Diagnosis, n (%)	
Hip and leg disorders	10 (26)
Sports injuries	7 (18)
Foot and ankle disorders	5 (13)
Scoliosis	4 (10)
Abdominal	3 (8)
Bone and soft tissue tumors	2 (5)
Neuromuscular	2 (5)
Other	6 (15)
Surgery, n (%)	
Orthopedic	
Hip and knee	15 (38)
Hardware removal or ablation	8 (21)
Foot and ankle	5 (13)
Spine	4 (10)
Other	4 (10)
General surgery	
Hernia repair	2 (5)
Urology	
Excision of penile cyst	1 (3)

Feasibility

Overview

Of the 39 patients, 6 (15%) used VR across the entire perioperative period, with 20 (51%) discontinuing VR in the preoperative waiting room, 6 (15%) inside the OR before induction, and 7 (18%) during induction (Figure 3). Transportation to the OR proceeded smoothly with no discontinuations. Most inductions were inhalational (37/39, 95%). Two (5%) IV inductions were attempted with VR, one of which was discontinued due to anxiety.

Figure 3. Timeline of VR intervention and reasons for discontinuation of VR. HCP: health care professional; OR: operating room; VR: virtual reality.



Facilitator: Receptiveness and Adaptability

Health care professionals appeared enthusiastic about the VR intervention, encouraging the patients during gameplay and offering implementation suggestions to the research team. On 6 occasions, the clinician adapted their preoperative evaluation in the waiting room to minimize interruption to the VR intervention by speaking to the parents or by discretely performing the task while the patient continued playing. After explaining the VR intervention and its implications for induction, all anesthesiologists attempted to integrate VR into their workflow. For example, the head of the OR table was raised, or extra pillows were placed under a patient's head for comfort while wearing the headset during induction. Only 2 (14%) of the 14 anesthesiologists in this study had prior VR experience.

Facilitator: Communication

Communication among patients, parents, and clinicians was maintained during the VR intervention. Clinicians explained what they were doing and notified the patient before important time points, such as leaving the waiting room for the OR. Clinicians and parents were actively involved in the VR intervention, asking children what they were seeing. Their involvement fostered a positive environment for patients to be immersed in VR. Patients could easily notify their clinicians about discomfort or desire to pause VR. Patients generally appreciated being informed of what was happening when using VR. One patient who was induced with VR reflected, "I would have liked to be notified when they were going to put on the induction mask. They didn't tell me! And I was surprised!" [Participant 29].

Barrier: Interruptions to the VR Intervention

Most patients (36/39, 92%) experienced at least one interruption during their entire VR intervention (from waiting room to induction), causing most patients (34/39, 87%) to remove their headset at least once. In the preoperative waiting room, the most common reason for interruption was preoperative assessments by clinicians (34/39, 87%), after which few patients (4/39, 10%) discontinued VR altogether. One mother stated, "[My daughter] was saying that she was so relaxed, and that VR helped her think about other things. And you could see that she was immersed in the game. And then, at one point, the effect was kind of lost because doctors came to see her" [Mother of participant 32].

Barrier: Duration of VR Use

The average total duration of the VR intervention was 23.1 (SD 24.4) minutes, ranging from 5 to 150 minutes. Most patients (37/39, 95%) had sufficient VR playtime in the preoperative waiting room to achieve immersive distraction for induction. However, due to frequent, unpredictable delays in the OR schedule, playtime in the waiting room was often extended for longer durations. Hence, some patients (15/39, 38%) took breaks in the middle of VR, and a few patients (4/39, 10%) became tired or bored, discontinuing VR altogether in the waiting room (Figure 3). One parent shared, "[…] at one point he stopped playing because it was always the same thing. After a while, it was enough" (parent of participant 10). In contrast, on one occasion, a patient [Participant 16] arrived late, leading clinicians to prioritize preoperative evaluations and reducing VR playtime to only 2 minutes before transfer to the OR. Nevertheless, this patient did not exhibit anxiety behaviors and had excellent compliance during induction (ICC=0).

Barrier: Induction With VR

Some challenges were noted in integrating VR into the intraoperative workflow. Among the 13 patients who attempted induction with VR, 6 (46%) inductions, including one IV induction, were completed with VR (Figure 3). VR was discontinued during 7 attempted VR inductions (7/13, 54%), including one IV induction, revealing challenges such as poor mask seal with the headset (n=3) and discomfort with supine positioning due to the headset structure (n=2; Figure 3). Achieving a good mask seal and ensuring a quick induction were prioritized by anesthesiologists over VR use, sometimes leading to VR discontinuation. Additionally, even when the

patient kept the headset during inhalational induction, one anesthesiologist explained that it was somewhat difficult to maintain a good mask seal. Additionally, 3 patients discontinued VR during the transfer from the stretcher to the OR table, and 2 temporarily paused VR. One patient explained that it felt like they were falling during the transfer from the stretcher to the OR bed with the headset.

Barrier: Technical Issues

VR-related technical issues, namely loss of audio (5/39, 13%), headset adjustments (5/39, 13%), and changes in the field of view when the patient changed orientation (4/39, 10%), were other sources of interruptions, at which point the researcher was able to quickly troubleshoot the issue, allowing for the patient to resume playing. On one occasion, the headset ran out of battery in the OR due to extended preoperative play time, causing a delay while the charger was retrieved, after which the patient continued to play.

Clinical Utility

Acceptability

VR was accepted by patients and parents. Almost all patients (38/39, 97%) were initially willing to use VR. Many patients looked forward to playing VR, asking the front desk personnel and nurses to commence upon arrival at the hospital. Parents supported the integration of VR in their child's care, asking them about the game and inquiring about hospital implementation efforts.

Satisfaction

Overall, all patients found VR fun (Table 2) and were happy to use VR perioperatively for distraction (Table 3), particularly in the preoperative waiting room. Several parents remarked their child was "gone in another world." One patient said, "I hope this never ends. Oh my god, this is so much fun" [Participant 20]. In contrast, adolescents primarily discontinued VR in the waiting room despite acknowledging the distraction it provided due to the game's repetitive and puerile nature. A few patients (3/39, 8%) opted to use other nonpharmacological means of distraction, such as their phone.

Table . Mean Graphic Rating Scale score for each item.

	Score, mean (SD)[a]
Pain	
Time spent thinking about pain	0 (0)
Unpleasant pain	0.1 (0.5)
Worst pain	0.4 (1.5)
Fun	7.8 (1.6)
Nausea	0 (0)

[a]Each item is rated on a 10 cm line, from 0 to 10. Along the line, descriptive markers "mild," "moderate" and "severe" are present.

Table . Patient perceptions of the clinical utility of the virtual reality intervention (n=17).

Scale and items	Responses, n (%)				Quotes
	1	2	3	4	
1=Not at all, 2=A little bit, 3=Some, and 4=A lot					
How much did the virtual reality game distract you during your medical procedure?	1 (5.9)	2 (11.8)	6 (35.3)	8 (47.1)	• "When they put me to sleep, I didn't even feel like I was being put to sleep. All I remember is having the mask on my face, being told to breathe, and then I was gone" [Participant 27].
How much did the virtual reality game help lower your pain during your medical procedure?	11 (64.7)	1 (5.9)	2 (11.8)	3 (17.6)	• "I didn't have pain to begin with." [Participant 32] • "Since I was focused on [VR], I would say that my pain [behind my knees] decreased." [Participant 1]
1=Very unlikely, 2=Unlikely, 3=Likely, and 4=Very likely					
Would you ask to play a virtual reality game for your next medical procedure?	0 (0)	0 (0)	9 (52.9)	8 (47.1)	• N/A
Would you recommend playing a virtual reality game to another patient like you?	0 (0)	0 (0)	5 (29.4)	12 (70.6)	• "This game would be good even for people my age who don't know much about video games. […] There are some more anxious, or who are having trouble coping, or for who it's their first surgery. VR would help them." [Participant 1]
1=Very unhappy, 2=Unhappy, 3=Happy, and 4=Very happy					

Scale and items	Responses, n (%)				Quotes
	1	2	3	4	
How happy were you with playing the virtual reality game during your medical procedure?	0 (0)	0 (0)	10 (58.8)	7 (41.2)	• "I hope this never ends! Can I have this for my birthday?" [Participant 20] • "Mom, I won 2585 points!" [Participant 44] • "It was okay for my age. It's not the best thing ever, but […] to distract me, it's pretty good." [Participant 27]

Ease of Use and Understanding

All patients, regardless of age and previous experience with VR or video games, rapidly understood how to play the game easily. Older patients were pre-emptively instructed by the researcher on troubleshooting certain technical issues, such as shifts in the field of view during position changes.

Recommendation of the VR Intervention

All patients would request and recommend VR if they or another patient needed another surgery under anesthesia (Table 3). One patient aged 17 years explained, "This game would be good even for people my age who don't know much about video games. […] There are some more anxious, or who are having trouble coping, or for whom it's their first surgery. VR would help them" [Participant 1].

Tolerability

Physical Adverse Events

All patients who used VR, regardless of duration, experienced no VR sickness at baseline, or during VR, as per the CSSQ (Table 4) and the GRS (Table 2). Two patients felt that their eyes were tired and took a break. One patient felt "a little bit dizzy," prompting him to take multiple short breaks in VR use in the waiting room. For the majority of children, the VR headset was comfortable. One patient found the headset "a little bit heavy on [her] head," which resolved when it was loosened. For another, the headset sometimes slid down his face. Discomfort was felt at the back of the head with 4 patients when asked to lay supine for induction with VR. One patient described a sensation of falling when being transferred from one bed to another with VR.

Table . Anxiety, pain, and virtual reality (VR) sickness: baseline versus during VR intervention.

Scale	Score, mean (SD)	
	Baseline (n=39)	During VR (n=17)
Faces Anxiety Scale[a]	1.5 (1.1)	0.7 (0.9)
Faces Pain Scale–Revised[b]	0.3 (0.9)	0.3 (1.2)
Child Simulator Sickness Questionnaire[c]		
Nausea	0 (0)	0 (0)
Oculomotor	0 (0)	0.02 (0.14)
Disorientation	0 (0)	0.02 (0.14)
Total	0 (0)	0.01 (0.13)

[a]The Faces Anxiety Scale for children is scored from 0 to 4, showing 5 faces with increasing levels of anxiety. A score of 0 means "no anxiety," a score of 4 means "extreme anxiety."

[b]The Faces Pain Scale–Revised is scored from 0 to 10, showing 6 faces with increasing pain. A score of 0 means "no pain," a score of 10 means "very much pain."

[c]A score of ≥3 of the Child Simulator Sickness Questionnaire indicates the presence of simulator sickness.

Emotional Adverse Events

The use of VR generated minimal adverse emotions. One patient [Participant 36], initially reluctant to use the headset due to a desire to see his surroundings, became immersed and distracted with reassurance.

Initial Clinical Efficacy

Anxiety

At baseline, the mean FAS score was 1.5 (SD 1.1), and many patients (24/39, 62%) demonstrated anxiety-related behaviors such as restlessness, crying, maintaining proximity to parents, and tense body language. Of the 39 patients, 2 (5%) were premedicated with midazolam prior to using VR due to particularly elevated anxiety (Table 1). During the VR intervention, the FAS score was 0.7 (SD 0.9) (Table 4), and some patients visibly relaxed as they became immersed, laughing and making exclamations about the game: “Wow! There’s lots of big balloons!” Patients expressed VR helped them cope: “It was fun, it made me stop thinking about the surgery completely” [Participant 27]. Parents echoed the sentiment, saying “[VR] definitely worked,” [Mother of participant 15]. Others viewed VR more pragmatically, describing VR as a tool that “allows [them] to pass time” [Participant 1] rather than management of anxiety. Patients who used VR during induction overall agreed that it distracted them, “I don’t remember what [health care professionals] were doing [during induction]” [Participant 25], “When they put me to sleep, I didn’t even feel like I was being put to sleep. All I remember is having the mask on my face, being told to breathe, and then I was gone” [Participant 27].

Spikes of anxiety, displayed as crying, verbal expressions of fear, tense body language, and withdrawing, were observed before transport to the OR and before initiating induction. Signs of poor immersion included decreased head movement and letting go of the controller. Due to increased anxiety, 1 patient discontinued VR in the waiting room before leaving for the OR, and 2 patients discontinued VR in the OR.

Induction Compliance

The majority of patients (32/39, 82%) had a perfect anesthetic induction (ICC=0), 5/39 (13%) patients had a suboptimal induction (ICC=1 to 4), whereas 2/39 (5%) patients had a poor induction (ICC>4). One patient [Participant 37] with a poor induction (ICC=9) was immersed in VR in the preoperative waiting room but, upon removing his headset in the OR, became rapidly anxious and agitated during induction.

Emergence Delirium

Upon recovery in the PACU, there was no report of emergency delirium. All patients scored zero on the PAED scale.

Pain

The majority of patients (25/39, 64%) had low baseline pain scores (Table 4) and did not perceive VR to help with pain management at all (Table 3). In this study, patients were not subject to painful interventions except for a preinduction IV insertion in two cases, one of which was aborted due to anxiety (Figure 3). In contrast, patients with baseline pain associated with their condition agreed that VR helped decrease it, “Because I was concentrated on other things, the pain decreased” [Participant 1].

Discussion

Principal Findings

Overall, while VR showed good clinical utility and tolerability, our study demonstrated feasibility challenges with the implementation of VR in the waiting room and induction. Importantly, there was a 50% discontinuation rate prior to arrival at the OR. In the waiting room, notable challenges included interruptions to VR by health care professionals in almost all patients and OR scheduling delays leading to unexpectedly long durations of VR use in the waiting room. This issue was compounded by the fact that DREAM was designed for short procedural distraction, limiting its suitability for longer durations of use when ORs were delayed. In contrast to other studies, our study contained patients in the adolescent age group, up to 18 years old, who compared to younger children may find VR less “novel” and have preferences for more complex games than DREAM. Additionally, the discontinuation rate was likely influenced by the study design and philosophy of care, which advocates for patients deciding if and how they would like to use VR, and to encourage the integration of their other coping strategies to relieve their anxiety, which is reflective of the real-world use of VR in practice. In other studies using VR at induction, premature discontinuation rates for VR were much lower, around 10% or less [19,20,23,24,26,36]. However, these studies introduced VR right before transport to the OR, whereas in this study, VR was often started in the waiting room, more than 5 minutes before transport to the OR, whether intentionally at the request of the patient or unintentionally due to delays.

Despite these feasibility barriers in the waiting room, VR demonstrated good clinical utility, as patients and parents reported high satisfaction and enjoyment with the VR intervention. They all recommended VR for others to use and desired to use VR again in the health care setting. FAS was relatively low at baseline in our study population and did not appear to change “during intervention”, though statistical significance and causality were not assessed for in this feasibility study. Previous VR studies in the waiting room have yielded positive results for anxiety reduction [7-16]. However, since these studies did preoperative OR tours by VR, establishing comparisons with our study and game may be difficult. Overall, waiting room VR games can be a valuable tool for temporary immersion and distraction; however, the suitability of VR becomes limited in cases of prolonged wait times.

The interpretation of the feasibility and use of VR during induction is limited by the high discontinuation rate in the waiting room and the consequently small sample of participants (6/39, 15%) induced with VR. Nevertheless, we noted that while VR was beneficial for some patients, for others, the distraction afforded by VR became limited as their anxiety increased in the OR and during induction. Thirteen inductions were attempted with VR in our study, half of which were completed without interruption of VR. Difficulties with mask fit and supine positioning were major feasibility barriers. The Pico Neo 3 headset used in this study has a hard piece of plastic at the back of the head, which rendered supine positioning uncomfortable for some patients. Troubleshooting included additional pillows

for padding and raising the head of the bed. As for troubles with the mask seal, the headset had to be propped up to allow access to the nose and mouth. In another study, anesthesiologists rotated the mask 180 degrees, allowing for a better fit with the headset at the expense of the mask seal [20].

To our knowledge, 2 RCTs have assessed the efficacy of VR during induction of anesthesia. Samnakay et al compared a VR video to a 2D video tablet, demonstrating similar efficacy between both technologies in reducing anxiety during induction. While children had higher satisfaction ratings with VR than with 2D tablets, anesthesiologists favored the 2D tablet over VR for inhalational inductions [20]. This is somewhat consistent with our findings, as we found induction with VR to be a technically challenging task that requires further optimization in technique and hardware, while satisfaction ratings remained high among patients. The similar efficacy of tablets and VR in their study [20], combined with the relative complexity of VR use during inhalational induction, argues against the use of VR during induction, though further evidence is needed to support this conclusion. Jung et al [19] demonstrated that a VR game, similar in mechanics to DREAM, during induction significantly decreased anxiety compared to the standard of care. In their study, only 1 out of 81 discontinued VR due to battery depletion, and 2 out of 81 discontinued VR during induction to see their parents [19]. This success, in contrast to our study, may be partly attributable to the use of a different headset (ie, Samsung Gear VR), in which the head strap is made of a softer, thinner material, not hindering supine positioning, and potentially to the health care professionals' experience levels with using VR.

Most (32/39, 82%) patients displayed perfect induction compliance as per the ICC, the interpretation of which becomes limited by the low number of patients wearing the VR headset at induction (6/39, 15%). Of note, one patient displayed poor induction compliance (ICC=9) only once the headset was removed in the OR. This is probably explained by their known prior history of high anxiety and poor induction compliance in the perioperative setting and the limited benefits that VR may offer certain patients. Furthermore, we observed that conflicting stimuli from the "real" environment, such as transfers from stretcher to OR bed and exposure to volatile anesthetics, can remove patients from their immersion. Interestingly, Samnakay et al reported children with VR had lower odds of having a perfect induction compared to children with tablets. Because VR hides the real-world environment, it is possible that real-world stimuli generate unintended surprises [20]. OR tours by VR in the waiting room improved induction compliance in two studies at the same institution [9,10], whereas they did not in two other studies [7,19]. In this study, a subset of patients preferred observing the OR environment during induction. This brings into consideration a potential advantage of augmented reality (AR) for them, in which a digital image is overlaid in the real world. The use of AR may significantly reduce anxiety in pediatric patients [37] and improve mask acceptance in children undergoing induction of GA compared to children induced without [38].

Clinical Implications

VR offers an innovative approach to help patients manage their anxiety before surgery under GA, but it is not a one-size-fits-all solution, and patients should be thoughtfully selected, especially considering the technical challenges encountered during induction. Importantly, the institution must be well organized for a coordinated approach to VR implementation. For minimal workflow and VR interruption, the intervention should ideally be started after completing all preoperative assessments. Ongoing collaboration and cooperation with all the health care providers should be elicited to minimize interruptions during gameplay. Indeed, a policy and procedure should detail when to alert cases of potential VR use with the health care team, especially anesthesiologists and respiratory therapists, such that they may adapt their approach and determine if VR is medically contraindicated. Prior to starting VR, the health care team should clarify with the patient when they would like to use VR, establish a communication plan, and determine whether the patient prefers being immersed or aware of their surroundings. However, they may always change their mind. It would be crucial to determine how VR for induction can be coordinated with expected and unexpected surgery delays. Depending on context, one health care professional should be responsible for administering and monitoring the VR intervention from the waiting room to the OR. Child life specialists, anesthesiologists, or respiratory therapists may be best equipped with that task as they are closely involved with the patient before and during induction. To render VR more compatible with induction, the health care team should opt for a headset that is not bulky, does not cover the patient's nose or mouth, and has no counterbalance weight at the back. Anesthesiologists should be aware that mask fit with the headset may be suboptimal, and that access to eyes is limited. The feasibility of VR may improve as the institution and its clinicians become increasingly experienced with its use.

Limitations

Due to the high discontinuation rate of VR preinduction, either by choice, surgical delays, or other circumstances, more data are needed to elucidate the feasibility of VR during anesthesia induction. Furthermore, many patients were lost to follow-up after their surgery; hence, self-reported postoperative outcomes such as anxiety and pain were incomplete. FAS "during VR" were obtained postintervention, in the PACU at the earliest, relying on the recall of the patients. Further, interviews were not conducted with all patients, potentially missing important insights for discontinuing VR use before induction. As this was a pilot feasibility VR2 trial, only descriptive statistics were performed, establishing no causal links. While this study included the perspectives of patients and parents, clinicians' perspectives were obtained via field notes, limiting our ability to offer a complete picture of VR benefits and limitations. The VR game DREAM was designed for young children during acute medical procedures, with some of the older patients expressing boredom leading to discontinuation. Finally, the significant portion of the playtime taking place in the waiting room may have influenced the discontinuation of VR use for the OR transfer and induction.

Future Research

Future studies aiming to investigate the use or implementation of VR in the perioperative setting should assess the feasibility of the intervention tailored to their organizational context. As mentioned previously, the feasibility of VR during induction of anesthesia could not be well assessed in this study due to the discontinuation rate and feasibility challenges that occurred prior to induction. Future studies should test the effectiveness of various games or software adapted to patient age, interests, and psychological needs. Further practice and research are needed to determine the conditions that would render VR compatible with anesthesia induction. More RCTs would be beneficial to truly assert the efficacy of VR in the perioperative period in comparison to other available technology, including 2D screens and augmented reality.

Conclusion

In the perioperative setting, from the waiting room until induction, VR may be a valuable tool for temporary distraction to help cope with this setting. While VR demonstrated clinical utility and tolerability, our study found, in the current state of VR implementation at our institution, important feasibility barriers in the waiting room and especially during the induction of anesthesia. Several considerations must be made to address the peculiarities of induction. This study contributes to the growing body of literature about VR in the perioperative process, elucidating important clinical challenges.

Acknowledgments

We would like to thank the health care professionals who have welcomed us and virtual reality into their practice for this study. The project was supported by the Réseau de Recherche en Santé Buccodentaire et Osseuse (RBSO) infrastructure grant and from our affiliated Shriners Temples Tunis and Rameses. AT is the recipient of a Junior 2 Research Scholar funded by the Fonds de Recherche du Québec (FRQ-S).

Conflicts of Interest

None declared.

References

1. Tas FQ, van Eijk CAM, Staals LM, Legerstee JS, Dierckx B. Virtual reality in pediatrics, effects on pain and anxiety: a systematic review and meta-analysis update. Paediatr Anaesth 2022 Dec;32(12):1292-1304. [doi: 10.1111/pan.14546] [Medline: 35993398]
2. Davidson AJ, Shrivastava PP, Jamsen K, et al. Risk factors for anxiety at induction of anesthesia in children: a prospective cohort study. Paediatr Anaesth 2006 Sep;16(9):919-927. [doi: 10.1111/j.1460-9592.2006.01904.x] [Medline: 16918652]
3. Perrott C, Lee CA, Griffiths S, Sury MRJ. Perioperative experiences of anesthesia reported by children and parents. Paediatr Anaesth 2018 Feb;28(2):149-156. [doi: 10.1111/pan.13300] [Medline: 29266767]
4. Mohkamkar MB, Farhoudi FM, Alam-Sahebpour AM, Mousavi SAM, Khani SP, Shahmohammadi SB. Postanesthetic emergence agitation in pediatric patients under general anesthesia. Iran J Pediatr 2014 Apr;24(2):184-190. [Medline: 25535538]
5. Kain ZN, Wang SM, Mayes LC, Caramico LA, Hofstadter MB. Distress during the induction of anesthesia and postoperative behavioral outcomes. Anesth Analg 1999 May;88(5):1042-1047. [doi: 10.1097/00000539-199905000-00013] [Medline: 10320165]
6. Fortier MA, Del Rosario AM, Rosenbaum A, Kain ZN. Beyond pain: predictors of postoperative maladaptive behavior change in children. Paediatr Anaesth 2010 May;20(5):445-453. [doi: 10.1111/j.1460-9592.2010.03281.x] [Medline: 20199608]
7. Park JW, Nahm FS, Kim JH, Jeon YT, Ryu JH, Han SH. The effect of mirroring display of virtual reality tour of the operating theatre on preoperative anxiety: a randomized controlled trial. IEEE J Biomed Health Inform 2019 Nov;23(6):2655-2660. [doi: 10.1109/JBHI.2019.2892485] [Medline: 30640637]
8. Ryu JH, Oh AY, Yoo HJ, Kim JH, Park JW, Han SH. The effect of an immersive virtual reality tour of the operating theater on emergence delirium in children undergoing general anesthesia: a randomized controlled trial. Paediatr Anaesth 2019 Jan;29(1):98-105. [doi: 10.1111/pan.13535] [Medline: 30365231]
9. Ryu JH, Park JW, Nahm FS, et al. The effect of gamification through a virtual reality on preoperative anxiety in pediatric patients undergoing general anesthesia: a prospective, randomized, and controlled trial. J Clin Med 2018 Sep 17;7(9):284. [doi: 10.3390/jcm7090284] [Medline: 30227602]
10. Ryu JH, Park SJ, Park JW, et al. Randomized clinical trial of immersive virtual reality tour of the operating theatre in children before anaesthesia. Br J Surg 2017 Nov;104(12):1628-1633. [doi: 10.1002/bjs.10684] [Medline: 28975600]
11. Wu Y, Chen J, Ma W, Guo L, Feng H. Virtual reality in preoperative preparation of children undergoing general anesthesia: a randomized controlled study. Anaesthesiologie 2022 Dec;71(Suppl 2):204-211. [doi: 10.1007/s00101-022-01177-w] [Medline: 35925196]

12. Eijlers R, Dierckx B, Staals LM, et al. Virtual reality exposure before elective day care surgery to reduce anxiety and pain in children: a randomised controlled trial. Eur J Anaesthesiol 2019 Oct;36(10):728-737. [doi: 10.1097/EJA.0000000000001059] [Medline: 31356373]
13. Dehghan F, Jalali R, Bashiri H. The effect of virtual reality technology on preoperative anxiety in children: a Solomon four-group randomized clinical trial. Perioper Med (Lond) 2019;8(5):5. [doi: 10.1186/s13741-019-0116-0] [Medline: 31171963]
14. Esposito C, Autorino G, Iervolino A, et al. Efficacy of a virtual reality program in pediatric surgery to reduce anxiety and distress symptoms in the preoperative phase: a prospective randomized clinical trial. J Laparoendosc Adv Surg Tech A 2022 Feb;32(2):197-203. [doi: 10.1089/lap.2021.0566] [Medline: 34962159]
15. Buyuk ET, Odabasoglu E, Uzsen H, Koyun M. The effect of virtual reality on children's anxiety, fear, and pain levels before circumcision. J Pediatr Urol 2021 Aug;17(4):567. [doi: 10.1016/j.jpurol.2021.04.008] [Medline: 34006462]
16. Castanys TF, Carrión AJ, Gómez FR, et al. Effects of virtual tour on perioperative pediatric anxiety. Paediatr Anaesth 2023 May;33(5):377-386. [doi: 10.1111/pan.14639] [Medline: 36700361]
17. Simonetti V, Tomietto M, Comparcini D, Vankova N, Marcelli S, Cicolini G. Effectiveness of virtual reality in the management of paediatric anxiety during the peri-operative period: a systematic review and meta-analysis. Int J Nurs Stud 2022 Jan;125:104115. [doi: 10.1016/j.ijnurstu.2021.104115] [Medline: 34781118]
18. Chen H, Chen L, Zhu C, Li S, Zhou J, Liu C. Immersive virtual reality versus video distraction for the management of emergence delirium in children: a randomized controlled study. J Perianesth Nurs 2025;40(2):318-325. [doi: 10.1016/j.jopan.2024.05.006] [Medline: 39140922]
19. Jung MJ, Libaw JS, Ma K, Whitlock EL, Feiner JR, Sinskey JL. Pediatric distraction on induction of anesthesia with virtual reality and perioperative anxiolysis: a randomized controlled trial. Anesth Analg 2021 Mar 1;132(3):798-806. [doi: 10.1213/ANE.0000000000005004] [Medline: 32618627]
20. Samnakay S, von Ungern-Sternberg BS, Evans D, et al. 3-Dimensional virtual reality versus 2-dimensional video for distraction during the induction of anesthesia in children to reduce anxiety: a randomized controlled trial. Anesth Analg 2024 Aug 23. [doi: 10.1213/ANE.0000000000007119] [Medline: 39178153]
21. Creswell JW, Fetters MD, Ivankova NV. Designing a mixed methods study in primary care. Ann Fam Med 2004;2(1):7-12. [doi: 10.1370/afm.104] [Medline: 15053277]
22. Birckhead B, Khalil C, Liu X, et al. Recommendations for methodology of virtual reality clinical trials in health care by an international working group: iterative study. JMIR Ment Health 2019 Jan 31;6(1):e11973. [doi: 10.2196/11973] [Medline: 30702436]
23. Addab S, Hamdy R, Le May S, Thorstad K, Tsimicalis A. The use of virtual reality during medical procedures in a pediatric orthopedic setting: a mixed-methods pilot feasibility study. Paediatr Neonatal Pain 2024 Sep;6(3):45-59. [doi: 10.1002/pne2.12078] [Medline: 39473832]
24. Gold JI, Annick ET, Lane AS, Ho K, Marty RT, Espinoza JC. "Doc McStuffins: Doctor for a Day" Virtual Reality (DocVR) for pediatric preoperative anxiety and satisfaction: pediatric medical technology feasibility study. J Med Internet Res 2021 Apr 19;23(4):e25504. [doi: 10.2196/25504] [Medline: 33730687]
25. Le May S, Hupin M, Khadra C, et al. Decreasing pain and fear in medical procedures with a pediatric population (DREAM): a pilot randomized within-subject trial. Pain Manag Nurs 2021 Apr;22(2):191-197. [doi: 10.1016/j.pmn.2020.10.002] [Medline: 33495093]
26. Osmanlliu E, Trottier ED, Bailey B, et al. Distraction in the emergency department using virtual reality for intravenous procedures in children to improve comfort (DEVINCI): a pilot pragmatic randomized controlled trial. CJEM 2021 Jan;23(1):94-102. [doi: 10.1007/s43678-020-00006-6] [Medline: 33683617]
27. Paperplane Therapeutics Virtual Reality. DREAM – virtual reality for pain & anxiety. YouTube. 2023 Jul 11. URL: https://www.youtube.com/watch?v=C0Y2dib7Hhg&ab_channel=PaperplaneTherapeutics [accessed 2025-04-10]
28. McKinley S, Coote K, Stein-Parbury J. Development and testing of a Faces Scale for the assessment of anxiety in critically ill patients. J Adv Nurs 2003 Jan;41(1):73-79. [doi: 10.1046/j.1365-2648.2003.02508.x] [Medline: 12519290]
29. Hicks CL, von Baeyer CL, Spafford PA, van Korlaar I, Goodenough B. The Faces Pain Scale-Revised: toward a common metric in pediatric pain measurement. Pain 2001 Aug;93(2):173-183. [doi: 10.1016/S0304-3959(01)00314-1] [Medline: 11427329]
30. Hoeft RM, Vogel J, Bowers CA. Kids get sick too: a proposed child simulator sickness questionnaire. Proc Hum Factors Ergon Soc Annu Meet 2003 Oct;47(20):2137-2141. [doi: 10.1177/154193120304702013]
31. Kain ZN, Mayes LC, Wang SM, Caramico LA, Hofstadter MB. Parental presence during induction of anesthesia versus sedative premedication: which intervention is more effective? Anesthesiology 1998 Nov;89(5):1147-1156. [doi: 10.1097/00000542-199811000-00015] [Medline: 9822003]
32. Tesler MD, Savedra MC, Holzemer WL, Wilkie DJ, Ward JA, Paul SM. The word-graphic rating scale as a measure of children's and adolescents' pain intensity. Res Nurs Health 1991 Oct;14(5):361-371. [doi: 10.1002/nur.4770140507] [Medline: 1891622]

33. Ford CG, Manegold EM, Randall CL, Aballay AM, Duncan CL. Assessing the feasibility of implementing low-cost virtual reality therapy during routine burn care. Burns 2018 Jun;44(4):886-895. [doi: 10.1016/j.burns.2017.11.020] [Medline: 29305105]
34. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
35. Farmer T, Robinson K, Elliott SJ, Eyles J. Developing and implementing a triangulation protocol for qualitative health research. Qual Health Res 2006 Mar;16(3):377-394. [doi: 10.1177/1049732305285708] [Medline: 16449687]
36. Mosadeghi S, Reid MW, Martinez B, Rosen BT, Spiegel BMR. Feasibility of an immersive virtual reality intervention for hospitalized patients: an observational cohort study. JMIR Ment Health 2016 Jun 27;3(2):e28. [doi: 10.2196/mental.5801] [Medline: 27349654]
37. Chamberland C, Bransi M, Boivin A, Jacques S, Gagnon J, Tremblay S. The effect of augmented reality on preoperative anxiety in children and adolescents: a randomized controlled trial. Paediatr Anaesth 2024 Feb;34(2):153-159. [doi: 10.1111/pan.14793] [Medline: 37925608]
38. Yun R, He EM, Zuniga M, et al. Augmented reality improves pediatric mask induction: a prospective, matched case-control study. J Patient Exp 2024;11. [doi: 10.1177/23743735241241146] [Medline: 38549806]

Abbreviations

AR: augmented reality
CSSQ: Child Simulator Sickness Questionnaire
FAS: Faces Anxiety Scale
FPS-R: Faces Pain Scale–Revised
GRS: Graphic Rating Scale
ICC: Induction Compliance Checklist
IV: intravenous
OR: operating room
PACU: postanesthesia care unit
PAED: Pediatric Anesthesia Emergence Delirium
RCT: randomized controlled trial
VR: virtual reality
VR-CORE: VR Clinical Outcomes Research Experts

Edited by S Pan; submitted 29.03.24; peer-reviewed by H Xiang, L Staals, M Görges; revised version received 16.02.25; accepted 09.03.25; published 16.05.25.

Please cite as:
Huang YT, Addab S, Bertolizio G, Hamdy R, Thorstad K, Tsimicalis A
Use of Virtual Reality in the Pediatric Perioperative Setting and for Induction of Anesthesia: Mixed Methods Pilot Feasibility Study
JMIR Perioper Med 2025;8:e58905
URL: https://periop.jmir.org/2025/1/e58905
doi:10.2196/58905

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.

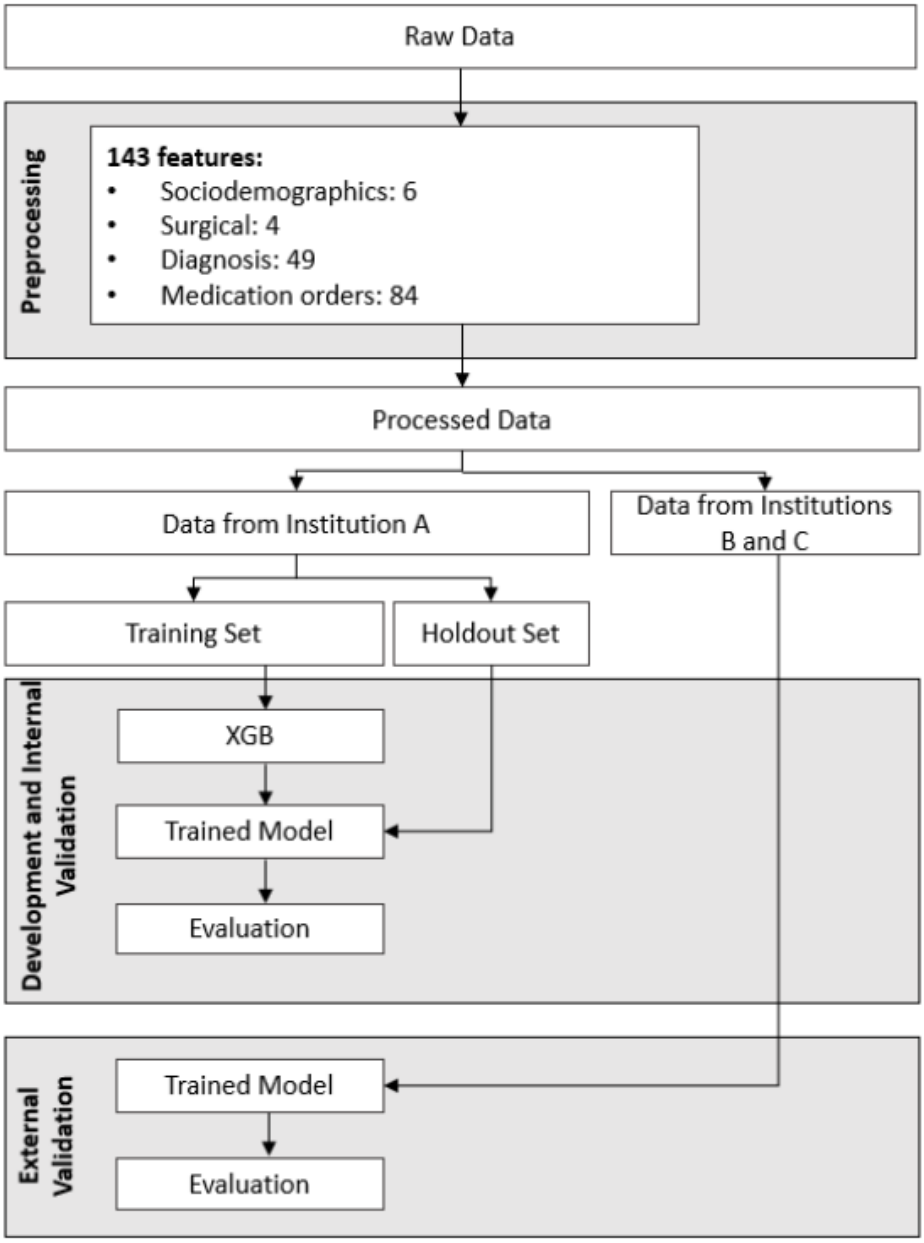
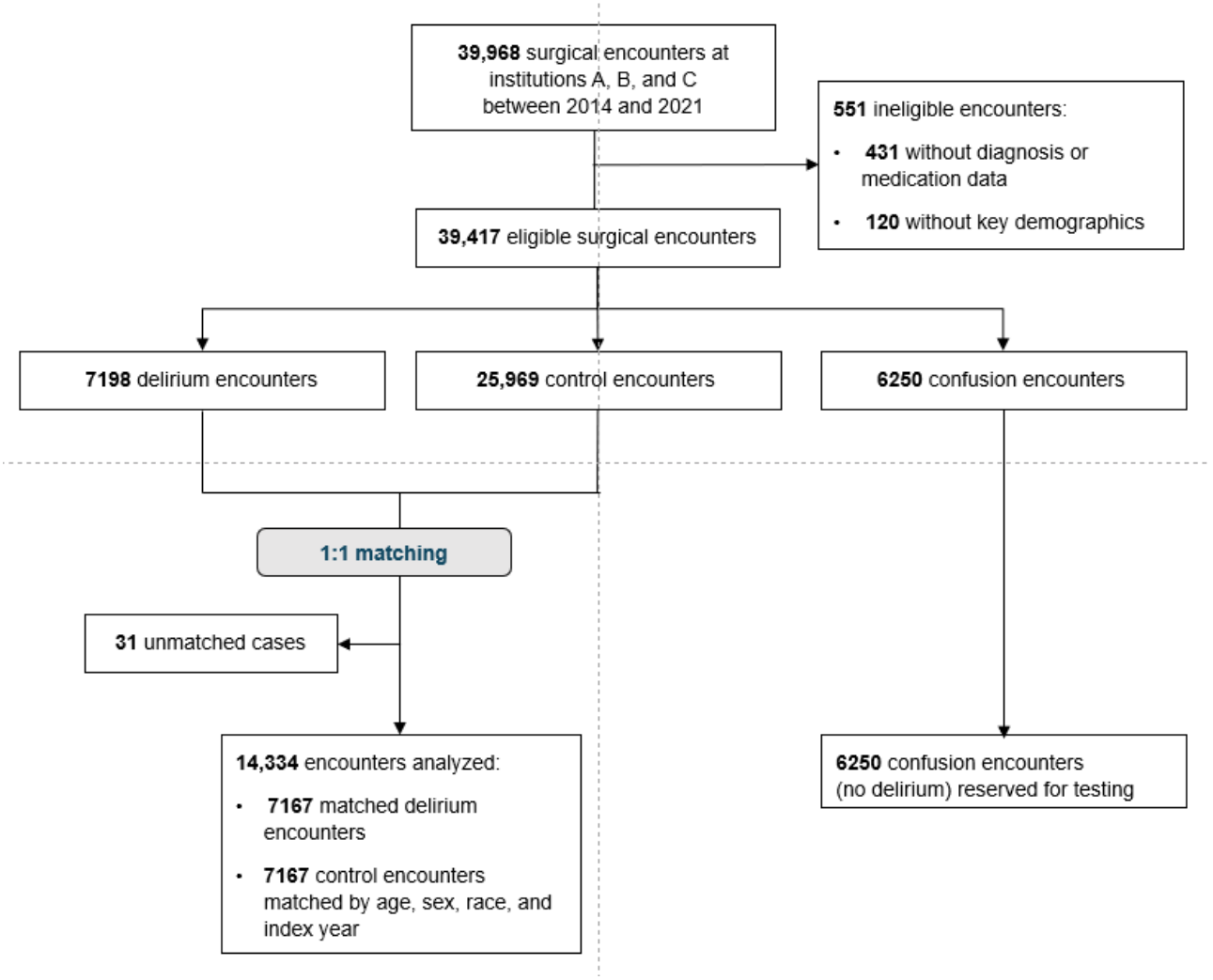


Figure 2. Patient inclusion flow diagram.



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:

Cost-Effectiveness of Day Surgery With Remote Patient Monitoring for Acute Cholecystitis: Economic Modeling Study

John Paul Kuwornu[1], PhD; David Brain[1], PhD; Kheng-Seong Ng[2], PhD; Amina Tariq[1], PhD; Melissa Baysari[3], PhD; Sundresan Naicker[1], PhD; Adeola Bamgboje-Ayodele[3,4], PhD; Adrian Boscolo[5], MBBS, MPH; Peter J Lee[6], MBBS, MPhil; Steven M McPhail[1,7], PhD

[1]Australian Centre for Health Services Innovation (AusHSI), School of Public Health and Social Work, Centre for Healthcare Transformation, Faculty of Health, Queensland University of Technology, Level 7, 88 Musk Avenue, Kelvin Grove QLD, Brisbane, Queensland, Australia

[2]Concord Institute of Academic Surgery, Concord Repatriation General Hospital, Concord, New South Wales, Australia

[3]Susan Wakil School of Nursing and Midwifery, Sydney School of Nursing, Faculty of Medicine and Health, The University of Sydney, Camperdown, New South Wales, Australia

[4]Discipline of Design, School of Architecture, Design and Planning, The University of Sydney, Camperdown, New South Wales, Australia

[5]Sydney Local Health District, Sydney, New South Wales, Australia

[6]Department of Colorectal Surgery, Royal Prince Alfred Hospital, Sydney, New South Wales, Australia

[7]Digital Health and Informatics Directorate, Metro South Health, Brisbane, Queensland, Australia

Corresponding Author:
Steven M McPhail, PhD
Australian Centre for Health Services Innovation (AusHSI), School of Public Health and Social Work, Centre for Healthcare Transformation, Faculty of Health, Queensland University of Technology, Level 7, 88 Musk Avenue, Kelvin Grove QLD, Brisbane, Queensland, Australia

Abstract

Background: Reducing the time to surgery for patients requiring cholecystectomy may lessen the risk of adverse outcomes. Dedicated day-surgery lists supported by out-of-hospital remote monitoring have been explored as a potential solution; however, the cost-effectiveness of such innovative care models remains largely unexplored.

Objective: This study presents a cost-effectiveness analysis comparing an acute day-surgery care model with remote patient monitoring to a conventional inpatient-centric care model for high-acuity cases of cholecystitis.

Methods: Post-surgical complications, effectiveness (measured by bed days saved and quality-adjusted life years [QALYs]), and health care costs associated with the two models of care were compared over a 1-year time horizon using a decision tree model. Health care costs were estimated from the Australian health care funder perspective and expressed in 2023 Australian dollars. Uncertainty was assessed using both deterministic and probabilistic sensitivity analyses.

Results: The acute day-surgery care model dominated the conventional inpatient-centric care model by saving a mean of 1.7 inpatient days per patient (3.2 days for the conventional model versus 1.5 days for the acute day-surgery model) and lowering net health care costs by a mean of AU $1,416 (US $935) per case over the 1-year time horizon. There was no meaningful difference in QALYs between the care models. These results remained robust in both deterministic and probabilistic sensitivity analyses.

Conclusions: An acute day-surgery care model with remote patient monitoring for individuals with acute cases of cholecystitis requiring cholecystectomy would likely free bed days and provide economic benefits to the health care system compared to inpatient-centric practice. Uncertainty in QALY estimates remains a limitation.

(JMIR Perioper Med 2025;8:e76807) doi:10.2196/76807

KEYWORDS

digital health; virtual care; economic evaluation; hospital; costs; laparoscopic surgery

Introduction

The effective management of surgical caseloads and theater resourcing can be challenging in high-demand hospital environments [1,2]. This is further complicated by critical care and surgical ward bed availability constraints [2,3]. International guidelines and local policies advocate for the prioritization of emergent surgical cases to preserve life and support patient safety [1-3]. Consequently, less urgent procedures may be postponed, which may be associated with an increased risk of adverse outcomes in some cases [1-5].

Cholecystectomy as an effective treatment for people with acute cholecystitis who have presented to hospital emergency

departments (EDs) is an important case in point [3-6]. Evidence from meta-analyses indicates that avoidable delay in time to surgery for patients requiring cholecystectomy may be associated with a greater risk of adverse outcomes and inefficient hospital resource use [4-6]. After presenting to the ED, patients with less urgent cases may initially be scheduled for next-day cholecystectomy procedures and remain as inpatients until the procedure and recovery are complete. However, for patients who have presented to hospitals that have large emergency caseloads, there is a significant and predictable risk that patients' cholecystectomy procedures will be de-prioritized relative to higher acuity cases, resulting in long lengths of stay in the hospital [3-6].

A potential solution that has been proposed is the dedicated day-surgery lists for low-risk cases, supported by out-of-hospital care, including the potential for remote monitoring through integrated virtual care or hospital-in-the-home care models [7-9]. The evidence around such novel care models is emerging, including meta-analyses that have indicated virtual wards and hospital-in-the-home care models can produce similar or potentially better outcomes for patients than conventional hospital inpatient care [7-9]. This innovative solution may have particular relevance for EDs with large critical and urgent caseloads, as well as multiple-theater arrays that are not fully used due to costs associated with labor resourcing for operating theaters. In these facilities, through allocation of additional resourcing for dedicated theater lists, there is potential for less urgent patients requiring cholecystectomy to return home after their initial presentation to the ED and present the following day for prompt planned day-surgery with the intention of returning home with remote patient monitoring and support, minimizing their overall inpatient stay.

While care models of this nature have the potential to reduce the length of stay and improve efficiency in resource use without negatively impacting patient outcomes [7-9], their cost-effectiveness remains largely unexplored. Cost-effectiveness is an important consideration for these cases, as a viable economic case must typically be made for the allocation of additional resourcing to establish dedicated day surgery lists for low-risk procedures of this nature. The aim of the present study was to estimate the cost-effectiveness of an acute day-surgery model of care, which may also be known as a same-day discharge model of care, with remote patient monitoring support in comparison to a conventional inpatient-centric model of care for acute cases of cholecystitis.

Methods

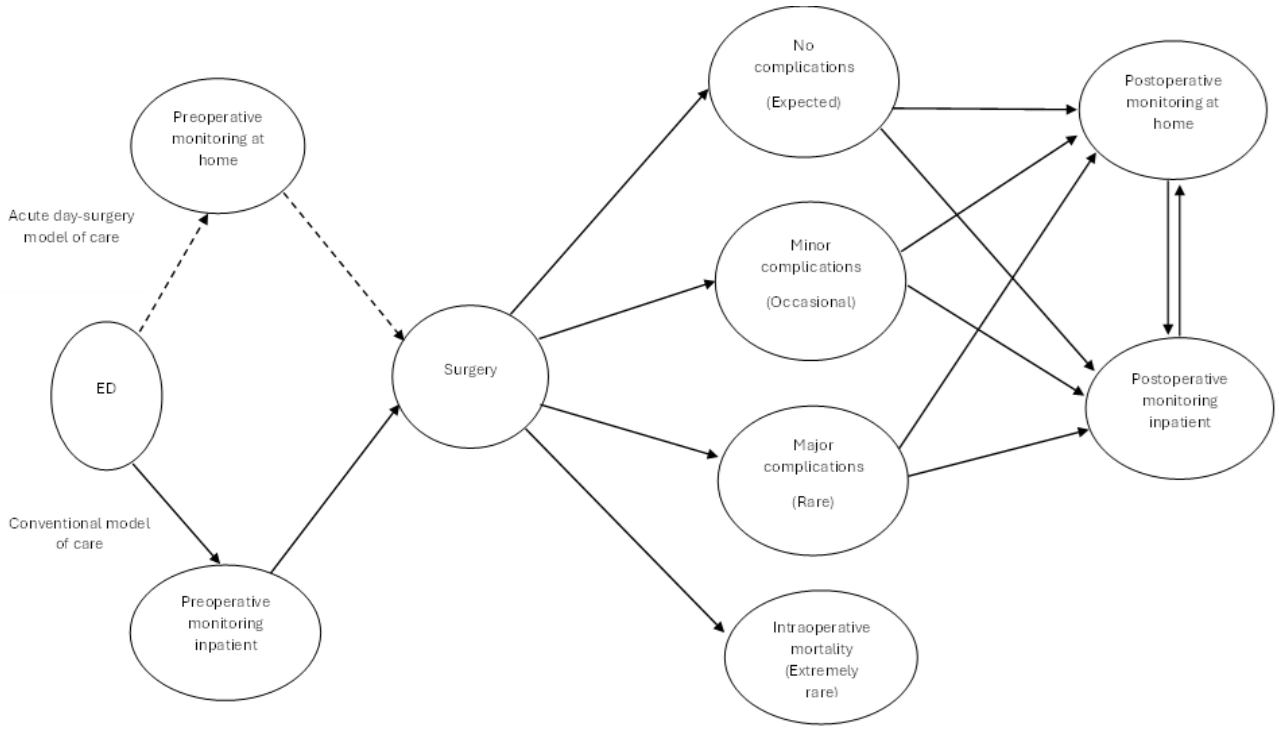
Scope of the Clinical Population

The population for the base case analysis comprised adult patients presenting to an ED with benign gallbladder disease for which cholecystectomy would typically be indicated on the same admission, and who were considered to have low surgical risk. Recently, based on evidence synthesis, Rickward et al [10] developed an "optimal" inclusion criterion for successful same-day cholecystectomies, which contributed to informing our base case. For example, the inclusion criterion considered patients younger than 65 years, American Society of Anaesthesiologists (ASA) physical status classification of 1 or 2, those with no prior upper abdominal surgeries, those with no or low risk for common bile duct stones, and those who had a responsible adult at home. The majority of biliary pathologies would include acute or chronic cholecystitis and intractable biliary colic [11]. The model assumed patients first presented to a hospital ED with a large emergency caseload within 24 hours of acute pain onset, and that diagnosis was confirmed by abdominal imaging. The following patient-type exclusions were considered out-of-scope for the modeling: previous receipt of cholecystectomy for the treatment of neoplasms, cholecystectomy concurrent with another procedure, unsuitable for minimally invasive surgical intervention or where open procedures would be planned from the outset [11].

Treatment Strategies

Our study compared two treatment strategies from the Australian health care funder perspective [1]: (1) a new acute day-surgery model of care with remote patient monitoring, and (2) a conventional inpatient-centric model of care. We adopted this perspective to inform future health care policy from the perspective of health care funders deciding whether to implement a dedicated surgical list for cholecystectomies. Under the conventional model of care, the patient is admitted as an inpatient after presentation to the ED with acute cholecystitis. The patient remains an inpatient until cholecystectomy is performed after allowing for potential delays due to emergency surgical cases consistent with hospitals that have large emergency caseloads. This contrasted with the new acute day-surgery model of care in which patients received initial assessment and surgical work-up in ED then returned home remaining under remote patient monitoring while a cholecystectomy is scheduled within 24 hours of the initial presentation on a dedicated 'same-day discharge' surgical list, on which other appropriate same-day acute cases (such as abscess drainages) would be operated on as same-day hospital admissions [12]. Under both treatment strategies, the surgical outcome determined the length of recovery time the patient spends post-surgery in the hospital. Under the new model of care, patients who experienced no complications post-surgery are placed on an expedited discharge protocol supported by remote patient monitoring, while those with post-surgical complications follow the usual care post-surgical inpatient protocol. In both care models, cholecystectomies could be performed laparoscopically or converted to open at the surgeon's discretion [13]. We focused on laparoscopic cholecystectomy and converted to open cholecystectomy procedures in this study because they account for approximately 95% of all cholecystectomies in Australia [10]. Figure 1 illustrates potential clinical pathways under the two models of care.

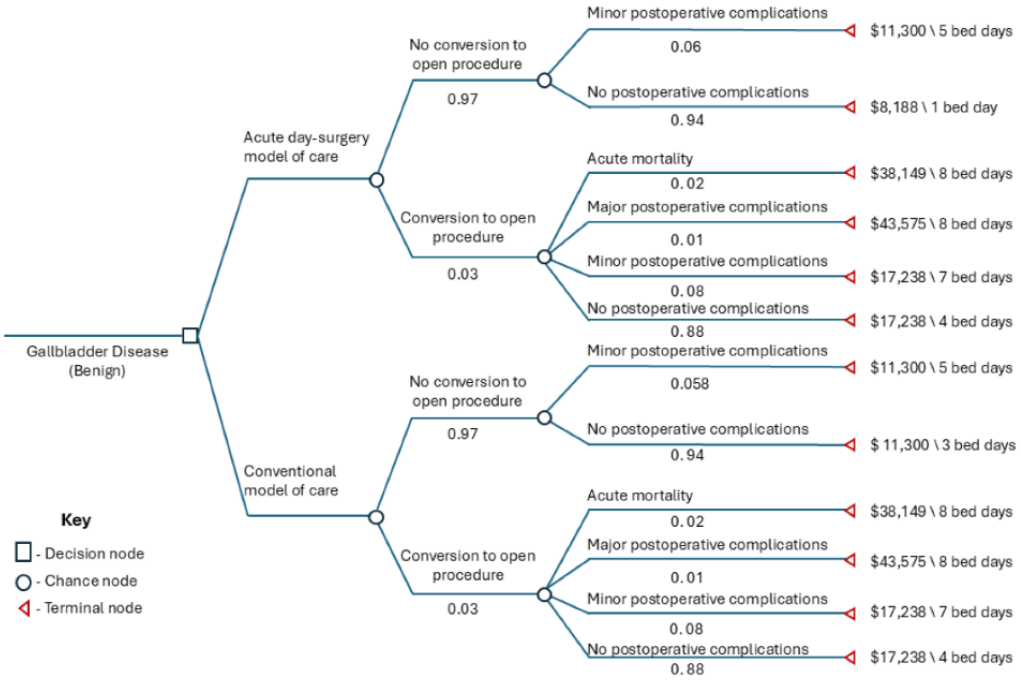
Figure 1. Patient clinical pathways and surgical outcome states under the new acute day-surgery model of care contrasted with the conventional model of care. ED: emergency department.



Decision Model

Using decision analysis software (TreeAge Pro, Williamstown, MA), a decision tree model (Figure 2) was created to analyze and compare the costs and outcomes of the two models of care over a 1-year time horizon, a time horizon frequently used in the cost-effectiveness analysis of cholecystectomy modalities [14-18]. We selected a decision tree model, as this approach was able to appropriately represent the associated costs and health outcomes, including potential complications, and is consistent with prior literature in the field [14-18]. Under each care model, the procedure was either completed laparoscopically or converted to open cholecystectomy. Patients then experienced either no postoperative complications, minor postoperative complications, major postoperative complications, or acute mortality [11]. The branch probabilities in the decision tree model, extracted from the existing literature [11], estimated the likelihood of a patient reaching these endpoints. Patients without postoperative complications and those with Clavien-Dindo grade 1/2 complications [19] were grouped together in the no postoperative complication group because they have similar lengths of stay [11]. Minor postoperative complications were considered surgery outcomes of Clavien-Dindo grade 3 complications [19], reserved for those requiring a procedure after a complication [11]. Major postoperative complications were those with any Clavien-Dindo grade 4 complications [19], described as any life-threatening complication requiring an intensive level of care [11]. Acute mortality was defined as the death of the patient within 30 days postoperatively (Figure 2).

Figure 2. The decision tree model structure. In decision trees, each node type represents a specific function in the decision-making process. A decision node (square) indicates a point where a choice between different strategies or actions is required. A chance node (circle) represents uncertainty, with outcomes determined by assigned probabilities. A terminal node (triangle) marks the end of a pathway, where model payoffs for patients reaching this outcome are summarized.



Model Variables: Probabilities

Probabilities for each decision tree branch used in the base case and sensitivity analyses are presented in Table 1. The probabilities of clinical events were extracted from published literature, with sources indicated in Table 1, including probabilities for converting to open cholecystectomies and minor or major postoperative complications.

Table . Model parameters extracted from the published literature.

Parameter	Base case estimate	Range [SD][a]	Source
Model parameters under the conventional model of care			
Decision tree branch probabilities			
Laparoscopic cholecystectomy without conversion to open	0.97	0.96 - 0.98 [0.18]	[14,20]
No complications	0.94	0.90 - 0.98 [0.18]	[14,20-22]
Minor complications	0.06	0.04 - 0.09 [0.84]	[14,20-22]
Laparoscopic cholecystectomy with conversion to open	0.03	0.02 - 0.05 [0.25]	[14,20]
No complications	0.88	0.84 - 0.94 [0.18]	[14,20]
Minor complications	0.08	0.04 - 0.16 [1.83]	[14,23]
Major complications	0.01	0.00 - 0.06 [0.94]	[14,23]
Acute mortality	0.02	0.01 - 0.08 [1.15]	[14,23]
Utilities			
Laparoscopic cholecystectomy without conversion to open			
No complications	0.98	0.93 - 0.99 [0.02]	[14,24]
Minor complications	0.97	0.92 - 0.99 [0.02]	[14,24]
Laparoscopic cholecystectomy with conversion to open			
No complications	0.87	0.82 - 0.92 [0.1]	[14,24]
Minor complications	0.85	0.78 - 0.90 [0.1]	[14,24]
Major complications	0.82	0.77 - 0.87 [0.1]	[14,24]
Acute mortality	0	0	
Bed days			
Laparoscopic cholecystectomy without conversion to open			
No complications	3	1 - 5	[14,20-22]
Minor complications	5	2 - 8	[14,20-22]
Laparoscopic cholecystectomy with conversion to open			
No complications	4	2 - 8	[14,22,25,26]
Minor complications	7	3 - 11	[14,22,25,26]
Major complications	8	4 - 15	[14,22,25,26]
Acute mortality	8	1 - 15	[14,22,25,26]
Health care costs (2023 AU $)[b]			
Laparoscopic cholecystectomy without conversion to open			
No complications	$11,300	$5,650 – $16,950 [$5,650]	[27]
Minor complications	$11,300	$5,650 – $16,950 [$5,650]	[27]
Laparoscopic cholecystectomy with conversion to open			
No complications	$17,238	$8,619 – $25,857 [$8,619]	[27]
Minor complications	$17,238	$8,619 – $25,857 [$8,619]	[27]
Major complications	$43,575	$21,787 – $65,362 [$21,787]	[27]

Parameter	Base case estimate	Range [SD][a]	Source
Acute mortality	$38,149	$19,074 – $57,223 [$19,074]	[27]
Model parameters under the acute day-surgery model of care			
Decision tree branch probabilities			
Laparoscopic cholecystectomy without conversion to open	0.97	0.96 - 0.98 [0.18]	[14,20]
No complications	0.94	0.90 - 0.98 [0.18]	[14,20-22]
Minor complications	0.06	0.04 - 0.09 [0.84]	[14,20-22]
Laparoscopic cholecystectomy with conversion to open	0.03	0.02 - 0.05 [0.25]	[14,20]
No complications	0.88	0.84 - 0.94 [0.18]	[14,20]
Minor complications	0.08	0.04 - 0.16 [1.83]	[14,23]
Major complications	0.01	0.00 - 0.06 [0.94]	[14,23]
Acute mortality	0.02	0.01 - 0.08 [1.15]	[14,23]
Utilities			
Laparoscopic cholecystectomy without conversion to open			
No complications	0.98	0.93 - 0.99 [0.02]	[14,24]
Minor complications	0.97	0.92 - 0.99 [0.02]	[14,24]
Laparoscopic cholecystectomy with conversion to open			
No complications	0.87	0.82 - 0.92 [0.1]	[14,24]
Minor complications	0.85	0.78 - 0.90 [0.1]	[14,24]
Major complications	0.82	0.77 - 0.87 [0.1]	[14,24]
Acute mortality	0	0	
Bed days			
Laparoscopic cholecystectomy without conversion to open			
No complications	1	1 - 2	[14,20-22]
Minor complications	5	2 - 8	[14,20-22]
Laparoscopic cholecystectomy with conversion to open			
No complications	4	2 - 8	[14,22,25,26]
Minor complications	7	3 - 11	[14,22,25,26]
Major complications	8	4 - 15	[14,22,25,26]
Acute mortality	8	1 - 15	[14,22,25,26]
Health care costs (2023 AU $)			
Laparoscopic cholecystectomy without conversion to open			
No complications	$8,188	$4,094 – $12,282 [$4,094]	[27]
Minor complications	$11,300	$5,650 – $16,950 [$5,650]	[27]
Laparoscopic cholecystectomy with conversion to open			
No complications	$17,238	$8,619 – $25,857 [$8,619]	[27]
Minor complications	$17,238	$8,619 – $25,857 [$8,619]	[27]

Parameter	Base case estimate	Range [SD][a]	Source
Major complications	$43,575	$21,787 – $65,362 [$21,787]	[27]
Acute mortality	$38,149	$19,074 – $57,223 [$19,074]	[27]
Remote patient monitoring	$1,556[c]	$500 – $5000 [$778]	[28]

[a]The range values were used in the one-way sensitivity analysis whilst the SD values were used in the probabilistic sensitivity analysis.

[b]As all costs were originally calculated in Australian dollars, a currency exchange rate of AU $1 = US $0.66 is applicable.

[c]We assumed that the remote patient monitoring cost component would not exceed $3112 per patient ($11,300-$8188), which was the cost reduction from converting conventional inpatient laparoscopic cholecystectomies to same-day procedures. The base case cost for remote patient monitoring was set at 50% of the cost reduction ($3,112 × 0.5 = $1556), equivalent to $518.60 per person-day for remotely monitoring a patient for 3 days.

Model Variables: Costs

The cost information for cholecystectomies was extracted from the Independent Hospital and Aged Care Pricing Authority [27]. This included information on laparoscopic cholecystectomies with varying degrees of complexity, ranging from minor to major, as well as for open cholecystectomies. The same cost for surgeries was applied for cases with no postoperative complications or minor postoperative complications only, but higher rates were applied for major complications. For laparoscopic cholecystectomies converted to open, we conservatively used the costs of open cholecystectomies. There was no published Australian cost information on cholecystectomies under the new model of care, so we used information from the literature to estimate these costs. Specifically, Manzia et al [28] found that conducting conventional inpatient laparoscopic cholecystectomies as same-day procedures reduced the cost per procedure by 1.38 times. We applied this ratio to local estimates for costs for laparoscopic cholecystectomies with no postoperative complications and parameterized this for the new same-day care model (Table 1). Our assumptions for other model parameterization were consistent with existing literature [7,9,29] that has indicated that hospital-in-the-home type care with remote patient monitoring is typically cheaper than the inpatient care alternative for uncomplicated surgical patients (Table 1). Remote patient monitoring costs included costs of equipping, establishing, and staffing the remote monitoring program for each patient case. Since all costs were originally calculated in Australian dollars, a currency exchange rate of AU $1 = US $0.66 is applicable.

Model Variables: Bed Days and Health-Related Quality of Life

An important motivation for considering laparoscopic cholecystectomies as same-day procedures stems from hospital operational efficiency innovation intended to free up inpatient bed capacity and potentially reduce waiting times for surgeries [7-9]. Consequently, our main outcome was inpatient bed days saved with the new model of care. Representation of health-related quality of life outcomes in this modeling was also considered important to clinicians and health care administrators contributing to this study as investigators or care model informants. It is common for modeling studies to apply literature-informed estimates for quality-adjusted life year (QALY)-related parameterization, and we were able to draw on literature to assign parameter values for health states represented in the model (Figure 1). However, it was challenging to accommodate potential between-care model differences in QALYs due to a lack of comparative effectiveness studies examining potential differences between the two care models represented in this study. On one hand, it might have been considered reasonable to expect that less time waiting in hospital for surgery, and fewer days in hospital overall to have a small beneficial effect on patients' health-related quality of life and thus assume favorable QALY parameterization in favor of the new care model. On the other hand, given that most patients will have relatively short lengths of stay in hospital in both care scenarios and that health-related quality of life is likely to primarily be influenced by factors unrelated to the care model, it may also be reasonable to assume that we would not detect a difference in QALY levels attributable to the new model of care. Therefore, in the absence of comparative effectiveness evidence for QALY effects, we adopted the more conservative approach and assumed the same QALY point estimates for both models of care (Table 1) for the present study, but parametrized considerable uncertainty to reflect and highlight the current lack of empirical QALY estimates. While this approach may be considered conservative and likely to produce indeterminate QALY findings, we thought it was appropriate to highlight this uncertainty and the importance of including patient-reported outcomes in future prospective comparative effectiveness studies in this field whether randomized trials or prospective quasi-experimental studies.

Cost-Effectiveness/Utility Analyses

The primary analysis ascertained the cost-effectiveness of the new acute day-surgery model of care compared to the conventional model of care, using the net incremental benefits approach with the outcome of the number of bed days saved. Net monetary benefit (NMB) estimates were informed by a willingness-to-pay (WTP) threshold derived from a recent local publication [30] originating from New South Wales, Australia. That study estimated the average inpatient cost for patients who received laparoscopic cholecystectomy was AU $3873 per day in 2019 [30], which, if inflated [31] to 2023 values, represents approximately AU $4522. We, therefore, adopted a slightly more conservative WTP threshold of AU $4000 per inpatient day. The model of care that generated the highest incremental net benefits was considered the most cost-effective. The secondary analysis used QALYs as the outcome, and the treatment strategy that generated the highest QALYs, without exceeding an ICER of AU $50,000/QALY was the most cost-effective. We selected the WTP threshold of AU

$50,000/QALY because it is the most used value in current health care cost-utility analysis in Australia. The conventional model of care was used as the reference group for all comparisons of costs and outcomes.

We followed the CHEERS guidelines for reporting economic evaluations (Checklist 1).

Sensitivity Analysis

We conducted several sensitivity analyses to explore how the base case results responded to uncertainties in the parameters. First, we conducted one-way sensitivity analyses to examine the effects of all model parameters on the base case results. As we found no specific cost information on postoperative laparoscopic cholecystectomy remote patient monitoring, we explored broad ranges for these costs. For example, one study [32] reported the cost of remote patient monitoring after same-day laparoscopic sleeve gastrectomy at US $3816 per day for all patients (n=20) monitored by the service, while another study [33] reported the average per person-day cost of US $24 for a remote patient monitoring program for post-discharge management of type 2 diabetes. Our one-way sensitivity analysis for remote patient monitoring cost ranged between AU $166 and AU $1,666 per person-day. We used a tornado diagram to summarize the results of the one-way sensitivity analysis.

Furthermore, using estimated distributions on each model parameter, we performed a probabilistic sensitivity analysis to assess uncertainty in the model results using a Monte Carlo simulation with 5000 samples. We simulated the decision tree branch probabilities and QALYs from beta distributions, bed days from the program evaluation review technique (PERT) distributions, and costs from gamma distributions (Table 1). We used the range of uncertainties extracted from the literature. In cases where no range of uncertainties was reported, we allowed parameters to vary by ±50% of the index value.

Ethical Considerations

This economic modeling study did not require review by the Queensland University of Technology (QUT) Human Research Ethics Committee. In accordance with QUT's ethics review policy and the National Statement on Ethical Conduct in Human Research (Section 5.1.22), research that involves only the use of existing, publicly available, non-identifiable data is exempt from ethics review. The data used in this study were obtained exclusively from published sources and contained no identifiable personal information.

Results

Base Case

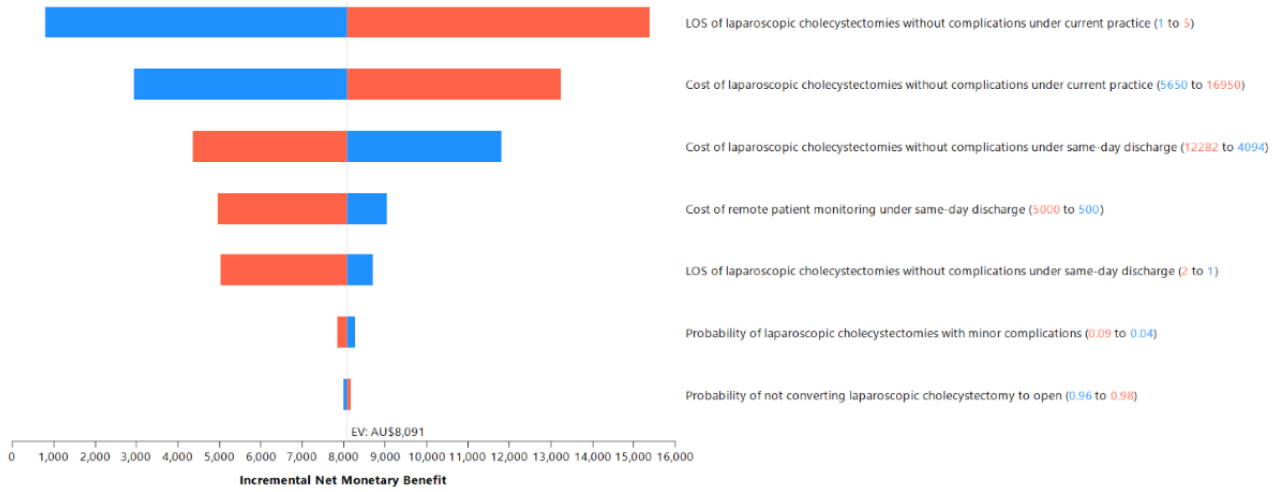
The base case results over a 1-year time horizon indicated that the new acute day-surgery model of care was dominant compared to the conventional model of care considering the main outcome of bed days saved. Specifically, the new model of care saved 1.7 days per patient (3.2 days for the conventional model of care vs 1.5 days for the acute day-surgery model of care) and lowered health care costs by AU $1416 per patient (AU $11,509 for the conventional model of care vs AU $10,093 for the acute day-surgery model of care). This translates to an incremental NMB of AU $8096 per case at the adopted WTP threshold.

Regarding the QALY outcome, the acute day-surgery model of care was marginally cost-effective compared to the conventional model of care. Both models were similar in effectiveness but had lower health care costs for the new care model (AU $11,509 for the conventional model of care vs AU $10,093 for the acute day-surgery model of care).

Sensitivity Analysis Results

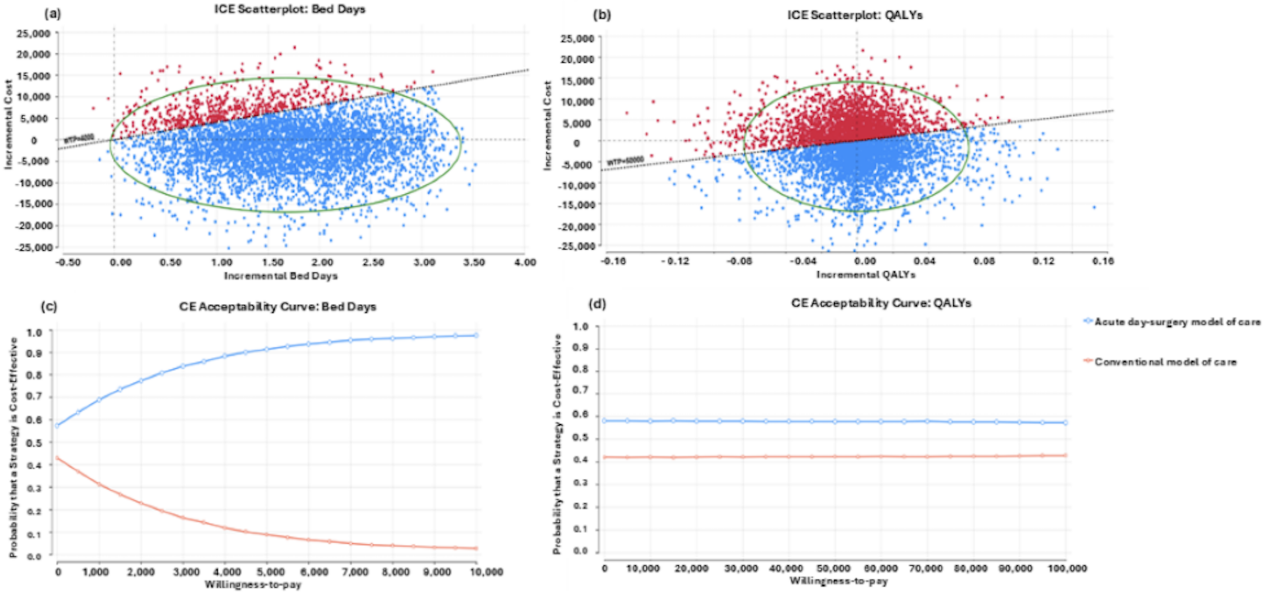
The deterministic one-way sensitivity analysis results are summarized in the tornado diagram in Figure 3. Findings indicated that the incremental NMB remained generally robust to the ranges of model input parameters explored. Of the 23 model parameters, only 7 had some sensitivity impact on the incremental NMB. The model was most sensitive to the length of stay parameterization for cholecystectomies without complications under the conventional model of care.

Figure 3. Tornado diagram (willingness-to-pay: AU $4,000). This tornado diagram ranks parameters by their influence on the net monetary benefit by varying each parameter across plausible ranges while holding others constant. The horizontal bars represent the range of results generated by varying each parameter from its minimum to maximum value. Parameters with the greatest impact appear at the top of the diagram, while those with minimal influence are shown at the bottom. The center line indicates the base-case outcome, and the left and right ends of each bar correspond to the outcome values at the lower and upper bounds of the parameter range, respectively. LOS: length of stay. Since all costs were originally calculated in Australian dollars, a currency exchange rate of AU $1 = US $0.66 is applicable.



The results of the probabilistic sensitivity analyses are presented in Figure 4. In large proportions of the 5000 re-samples, the new acute day-surgery model of care was either dominant or cost-effective (Figure 4A). There was an 89% probability that the new same-day care model would be cost-effective compared to the conventional model of care at a WTP threshold of AU $4,000 per acute surgical bed day saved (Figure 4C). Similarly, regarding the QALY outcome, the probabilistic sensitivity analysis results showed that the new acute day-surgery model of care was more likely to be cost-effective compared to the conventional model of care (Figure 4B), with a 57% probability of being cost-effective at a WTP threshold of AU $50,000 per QALY.

Figure 4. Results of the probabilistic sensitivity analysis presented in the cost-effectiveness planes for bed days (A) and quality-adjusted life years (QALYs) (B), as well as cost-effectiveness acceptability curves for bed days (C) and QALYs (D).



Discussion

Principal Results

This study presents the first cost-effectiveness analysis comparing a dedicated acute day-surgery model of care supported by virtual care with remote patient monitoring to a conventional inpatient-centric model of care for acute cases of cholecystitis. The results indicate that the dedicated acute day-surgery care model with remote patient monitoring was likely to be dominant (ie, saves bed days and incurs lower health care costs). Regarding QALYs, the new same-day discharge care model may be considered marginally cost-effective compared to the conventional inpatient-centric practice, which reflected our conservative approach of not assuming potential

improvements in QALYs that may be associated with less time in hospital. Both the one-way and probabilistic sensitivity analyses indicated that findings remained robust across potential model parameter ranges.

Comparison With Prior Work

Our study is aligned with the latest developments in the field. A systematic review of discordant meta-analysis [4] recommended early laparoscopic cholecystectomy over delayed laparoscopic cholecystectomy. One strategy often suggested to expedite surgery for patients with acute biliary disease is to dedicate theater resources and surgical expertise to same-day discharge protocols for laparoscopic cholecystectomies [8]. A single hospital visit pathway for day-case laparoscopic cholecystectomy has been trialed and found to be feasible, safe, and acceptable for patients with symptomatic gallstone disease [34]. This is consistent with other literature [28] that established ambulatory laparoscopic cholecystectomy as both safe and cost-effective. A recent systematic review [10] has also reported an "optimal" inclusion criterion for successful same-day cholecystectomies in Australia. It is also noteworthy that there have been previous favorable trials of protocols offering patients same-day laparoscopic sleeve gastrectomy supported by remote patient monitoring [32,35], and an acute day-only surgery program for abscess drainage [12].

Recent systematic reviews [7,9,29] including studies from other clinical populations have concluded virtual wards, hospital-in-the-home, and other remote patient monitoring interventions had positive impacts on patient safety, adherence, patients' mobility, functional statuses, and cost-related outcomes. However, investigations of the impact of remote patient monitoring on quality-of-life indicators remain inconclusive [7,9,29]. Other studies examining safety and feasibility of assessing and counseling patients for laparoscopic cholecystectomy remotely without a physical encounter [36] support the use of remote follow-up [37,38]. Similarly, a virtual clinic for post-operative patients who underwent laparoscopic cholecystectomy improved clinic efficiency [39].

While the remote patient monitoring evidence base is growing, evidence regarding the cost and cost-effectiveness implications of these innovations is lagging. A scoping review [40] considering economic impacts concluded that telehealth provides positive patient benefits and improves productivity for many services, but does not routinely reduce costs for health care systems. A systematic review of economic evaluations of remote patient monitoring interventions for chronic diseases found that remote patient monitoring interventions were highly cost-effective for hypertension, differed according to disease severity for chronic obstructive pulmonary disease and heart failure, and had limited economic evidence among patients with diabetes [41]. Our results extend this evidence base by indicating that bundling acute day surgeries for cholecystitis with virtual care, including remote monitoring, is likely to be cost-effective.

Limitations

This study has several limitations. Comparative-effectiveness evidence for the impact of remote patient monitoring use on QALYs in this patient group is lacking, and it may be important to update these modeling results as new evidence is reported. Randomized trials examining the impact of novel care models, including remote-patient monitoring, would be informative in this regard. Furthermore, we did not account for procedural complications such as incisional hernia because they are relatively rare, and we considered them unlikely to impact findings [14]. Similarly, the lack of granular public domain cost information for pre- and post-operative remote patient monitoring meant that we applied a wide range of potential costs, but fortunately, the sensitivity analyses indicated results were robust to this uncertainty.

Conclusions

The acute day-surgery model of care supported by remote patient monitoring was dominant. It is likely to save bed days and incur lower health care costs compared to the conventional inpatient-centric care model for acute cholecystitis cases.

Acknowledgments

The authors thank Owen Hutchings and the RPA virtual team for project support and advice. This study was supported by Digital Health CRC Limited (DHCRC). DHCRC is funded under the Australian Commonwealth's Cooperative Research Centers (CRC) Program. The funder was not involved in study design, data collection, data analysis, preparation and publication decisions of this manuscript, and the authors had complete access to the study data that support the publication.

Data Availability

The data used in this study were extracted from published sources.

Conflicts of Interest

None declared.

Checklist 1
CHEERS 2022 Checklist
[DOCX File, 26 KB - periop_v8i1e76807_app1.docx]

References

1. Pisano M, Allievi N, Gurusamy K, et al. 2020 world society of emergency surgery updated guidelines for the diagnosis and treatment of acute calculus cholecystitis. World J Emerg Surg 2020 Nov 5;15(1):61. [doi: 10.1186/s13017-020-00336-x] [Medline: 33153472]
2. Hassanzadeh H, Boyle J, Khanna S, Biki B, Syed F. Daily surgery caseload prediction: towards improving operating theatre efficiency. BMC Med Inform Decis Mak 2022 Jun 7;22(1):151. [doi: 10.1186/s12911-022-01893-8] [Medline: 35672729]
3. McCabe R, Schmit N, Christen P, et al. Adapting hospital capacity to meet changing demands during the COVID-19 pandemic. BMC Med 2020 Oct 16;18(1):329. [doi: 10.1186/s12916-020-01781-w] [Medline: 33066777]
4. Song GM, Bian W, Zeng XT, Zhou JG, Luo YQ, Tian X. Laparoscopic cholecystectomy for acute cholecystitis: early or delayed?: Evidence from a systematic review of discordant meta-analyses. Medicine (Baltimore) 2016 Jun;95(23):e3835. [doi: 10.1097/MD.0000000000003835] [Medline: 27281088]
5. Wu XD, Tian X, Liu MM, Wu L, Zhao S, Zhao L. Meta-analysis comparing early versus delayed laparoscopic cholecystectomy for acute cholecystitis. Br J Surg 2015 Oct;102(11):1302-1313. [doi: 10.1002/bjs.9886] [Medline: 26265548]
6. Abe T, Kobayashi T, Kuroda S, et al. Multicenter analysis of the efficacy of early cholecystectomy and preoperative cholecystostomy for severe acute cholecystitis: a retrospective study of data from the multi-institutional database of the Hiroshima Surgical Study Group of Clinical Oncology. BMC Gastroenterol 2024 Oct 1;24(1):338. [doi: 10.1186/s12876-024-03420-7] [Medline: 39354370]
7. Shi C, Dumville J, Rubinstein F, et al. Inpatient-level care at home delivered by virtual wards and hospital at home: a systematic review and meta-analysis of complex interventions and their components. BMC Med 2024 Apr 2;22(1):145. [doi: 10.1186/s12916-024-03312-3] [Medline: 38561754]
8. Jamdar S, Chandrabalan VV, Obeidallah R, Stathakis P, Siriwardena AK, Sheen AJ. The impact of a dedicated “hot list” on the in-patient management of patients with acute gallstone-related disease. Front Surg 2021;8:643077. [doi: 10.3389/fsurg.2021.643077] [Medline: 34055866]
9. McGillion MH, Parlow J, Borges FK, et al. Post-discharge after surgery Virtual Care with Remote Automated Monitoring-1 (PVC-RAM-1) technology versus standard care: randomised controlled trial. BMJ 2021 Sep 30;374:n2209. [doi: 10.1136/bmj.n2209] [Medline: 34593374]
10. Rickward J, Hameed I, Ho S, Wijeratne S. Day case laparoscopic cholecystectomy: a review of patient selection factors and identification of potential barriers to same-day discharge. ANZ J Surg 2024 Dec;94(12):2119-2127. [doi: 10.1111/ans.19241] [Medline: 39380458]
11. Singh A, Panse NS, Prasath V, Arjani S, Chokshi RJ. Cost-effectiveness analysis of robotic cholecystectomy in the treatment of benign gallbladder disease. Surgery 2023 Jun;173(6):1323-1328. [doi: 10.1016/j.surg.2023.01.017] [Medline: 36914510]
12. Yang PF, Builth-Snoad L, Ng KS, et al. Optimizing theatre utilization for abscess drainage: going beyond priority categories. ANZ J Surg 2024 Apr;94(4):648-654. [doi: 10.1111/ans.18919] [Medline: 38426392]
13. Breitenstein S, Nocito A, Puhan M, Held U, Weber M, Clavien PA. Robotic-assisted versus laparoscopic cholecystectomy: outcome and cost analyses of a case-matched control study. Ann Surg 2008 Jun;247(6):987-993. [doi: 10.1097/SLA.0b013e318172501f] [Medline: 18520226]
14. Singh A, Kaur M, Swaminathan C, Siby J, Singh KK, Sajid MS. Laparoscopic versus robotic cholecystectomy: a systematic review with meta-analysis to differentiate between postoperative outcomes and cost-effectiveness. Transl Gastroenterol Hepatol 2024;9:3. [doi: 10.21037/tgh-23-56] [Medline: 38317744]
15. Sugiura K, Suzuki K, Umeyama T, Omagari K, Hashimoto T, Tamura A. Cost-effectiveness analysis of initial nonoperative management versus emergency laparoscopic appendectomy for acute complicated appendicitis. BMC Health Serv Res 2020 Nov 9;20(1):1019. [doi: 10.1186/s12913-020-05839-6] [Medline: 33167993]
16. Ali A, Mobarak Z, Al-Jumaily M, et al. Cost-utility analysis of antibiotic therapy versus appendicectomy for acute uncomplicated appendicitis. Int J Environ Res Public Health 2021 Aug 11;18(16):8473. [doi: 10.3390/ijerph18168473] [Medline: 34444222]
17. Wu JX, Sacks GD, Dawes AJ, DeUgarte D, Lee SL. The cost-effectiveness of nonoperative management versus laparoscopic appendectomy for the treatment of acute, uncomplicated appendicitis in children. J Pediatr Surg 2017 Jul;52(7):1135-1140. [doi: 10.1016/j.jpedsurg.2016.10.009] [Medline: 27836368]
18. Guevara-Cuellar CA, Rengifo-Mosquera MP, Parody-Rúa E. Cost-effectiveness analysis of nonoperative management versus open and laparoscopic surgery for uncomplicated acute appendicitis in Colombia. Cost Eff Resour Alloc 2021 Jun 10;19(1):34. [doi: 10.1186/s12962-021-00288-2] [Medline: 34112179]
19. Clavien PA, Barkun J, de Oliveira ML, et al. The Clavien-Dindo classification of surgical complications: five-year experience. Ann Surg 2009 Aug;250(2):187-196. [doi: 10.1097/SLA.0b013e3181b13ca2] [Medline: 19638912]
20. Terho PM, Leppäniemi AK, Mentula PJ. Laparoscopic cholecystectomy for acute calculous cholecystitis: a retrospective study assessing risk factors for conversion and complications. World J Emerg Surg 2016;11:54. [doi: 10.1186/s13017-016-0111-4] [Medline: 27891173]
21. Han C, Shan X, Yao L, et al. Robotic-assisted versus laparoscopic cholecystectomy for benign gallbladder diseases: a systematic review and meta-analysis. Surg Endosc 2018 Nov;32(11):4377-4392. [doi: 10.1007/s00464-018-6295-9] [Medline: 29956028]

22. Halilovic H, Hasukic S, Matovic E, Imamovic G. Rate of complications and conversions after laparoscopic and open cholecystectomy. Med Arh 2011;65(6):336-338. [doi: 10.5455/medarh.2011.65.336-338] [Medline: 22299293]
23. Livingston EH, Rege RV. A nationwide study of conversion from laparoscopic to open cholecystectomy. Am J Surg 2004 Sep;188(3):205-211. [doi: 10.1016/j.amjsurg.2004.06.013] [Medline: 15450821]
24. Morris S, Gurusamy KS, Patel N, Davidson BR. Cost-effectiveness of early laparoscopic cholecystectomy for mild acute gallstone pancreatitis. Br J Surg 2014 Jun;101(7):828-835. [doi: 10.1002/bjs.9501] [Medline: 24756933]
25. Soffer D, Blackbourne LH, Schulman CI, et al. Is there an optimal time for laparoscopic cholecystectomy in acute cholecystitis? Surg Endosc 2007 May;21(5):805-809. [doi: 10.1007/s00464-006-9019-5] [Medline: 17180290]
26. Cao J, Liu B, Li X, et al. Analysis of delayed discharge after day-surgery laparoscopic cholecystectomy. Int J Surg 2017 Apr;40:33-37. [doi: 10.1016/j.ijsu.2017.02.055] [Medline: 28235668]
27. Independent hospital and aged care pricing authority. Costs of acute admitted patients in public hospitals. 2024. URL: https://www.ihacpa.gov.au/admitted-acute-care [accessed 2025-10-06]
28. Manzia TM, Quaranta C, Filingeri V, et al. Feasibility and cost effectiveness of ambulatory laparoscopic cholecystectomy. A retrospective cohort study. Ann Med Surg 2020 Jul;55:56-61. [doi: 10.1016/j.amsu.2020.04.036] [Medline: 32461804]
29. Tan SY, Sumner J, Wang Y, Wenjun Yip A. A systematic review of the impacts of remote patient monitoring (RPM) interventions on safety, adherence, quality-of-life and cost-related outcomes. NPJ Digit Med 2024 Jul 18;7(1):192. [doi: 10.1038/s41746-024-01182-w] [Medline: 39025937]
30. Merret N, Celan R, Smith K, Treloar B. Agency for clinical invitation. Same day laparoscopic cholecystectomy: reducing surgical procedure length of stay. 2024. URL: https://aci.health.nsw.gov.au/ie/projects/same-day-laparoscopic-cholecystectomy [accessed 2024-11-27]
31. Inflation calculator. Reserve Bank of Australia. 2024. URL: https://www.rba.gov.au/calculator/annualDecimal.html [accessed 2024-12-11]
32. Schaffner TJ, Wilkes M, Laverty R, et al. Remote patient monitoring to facilitate same-day discharge after laparoscopic sleeve gastrectomy: a pilot evaluation. Surg Obes Relat Dis 2023 Sep;19(9):1067-1074. [doi: 10.1016/j.soard.2023.02.028] [Medline: 37105773]
33. Michaud TL, Hill JL, Estabrooks PA, Su D. Cost analysis of a remote patient monitoring programme for post-discharge patients with type 2 diabetes. J Telemed Telecare 2023 Jul;29(6):417-425. [doi: 10.1177/1357633X20985393] [Medline: 33497310]
34. Curtis NJ, Robinson PD, Carty NJ. Single hospital visit elective day-case laparoscopic cholecystectomy without prior outpatient attendance. Surg Endosc 2017 Sep;31(9):3574-3580. [doi: 10.1007/s00464-016-5387-7] [Medline: 28127716]
35. Landreneau JP, Agarwal D, Witkowski E, et al. Safety and cost of performing laparoscopic sleeve gastrectomy with same day discharge at a large academic hospital. Surg Endosc 2024 Apr;38(4):2212-2218. [doi: 10.1007/s00464-024-10673-6] [Medline: 38379004]
36. Urbonas T, Lakha AS, King E, et al. The safety of telemedicine clinics as an alternative to in-person preoperative assessment for elective laparoscopic cholecystectomy in patients with benign gallbladder disease: a retrospective cohort study. Patient Saf Surg 2023 Aug 29;17(1):23. [doi: 10.1186/s13037-023-00368-7] [Medline: 37644474]
37. Daliya P, Carvell J, Rozentals J, Lobo DN, Parsons SL. Digital follow-up after elective laparoscopic cholecystectomy: a feasibility study. World J Surg 2022 Nov;46(11):2648-2658. [doi: 10.1007/s00268-022-06684-w] [Medline: 35953737]
38. Abbitt D, Choy K, Castle R, et al. Telehealth follow-up after cholecystectomy is safe in veterans. Surg Endosc 2023 Apr;37(4):3201-3207. [doi: 10.1007/s00464-022-09501-6] [Medline: 35974252]
39. Parnell K, Kuhlenschmidt K, Madni D, et al. Using telemedicine on an acute care surgery service: improving clinic efficiency and access to care. Surg Endosc 2021 Oct;35(10):5760-5765. [doi: 10.1007/s00464-020-08055-9] [Medline: 33048233]
40. Snoswell CL, Taylor ML, Comans TA, Smith AC, Gray LC, Caffery LJ. Determining if telehealth can reduce health system costs: scoping review. J Med Internet Res 2020 Oct 19;22(10):e17298. [doi: 10.2196/17298] [Medline: 33074157]
41. De Guzman KR, Snoswell CL, Taylor ML, Gray LC, Caffery LJ. Economic evaluations of remote patient monitoring for chronic disease: a systematic review. Value Health 2022 Jun;25(6):897-913. [doi: 10.1016/j.jval.2021.12.001] [Medline: 35667780]

Abbreviations

ASA: American Society of Anaesthesiologists
ED: emergency department
NMB: net monetary benefit
QALY: quality-adjusted life year
QUT: Queensland University of Technology
WTP: willingness-to-pay

Edited by S Pan; submitted 30.04.25; peer-reviewed by C Gissel, K Sangha; revised version received 24.08.25; accepted 24.09.25; published 20.10.25.

Please cite as:
Kuwornu JP, Brain D, Ng KS, Tariq A, Baysari M, Naicker S, Bamgboje-Ayodele A, Boscolo A, Lee PJ, McPhail SM
Cost-Effectiveness of Day Surgery With Remote Patient Monitoring for Acute Cholecystitis: Economic Modeling Study
JMIR Perioper Med 2025;8:e76807
URL: https://periop.jmir.org/2025/1/e76807
doi:10.2196/76807

Social Media Influence on Surgeon Selection Among Iranian Maxillofacial Patients: Cross-Sectional Survey Study

Mehdi Abrishami[1]; Milad Bayat[2]; Elham Sheykhi Some[3], MSc

[1]Department of Oral & Maxillofacial Surgery, Faculty of Dentistry, Isf. C, Islamic Azad University, Isfahan, Iran

[2]School of Dentistry, Islamic Azad University, Isfahan, Isfahan, Iran

[3]School of Health Management and Information Sciences, Iran University of Medical Sciences, No. 6, Shahid Yasami St, Vali-e-Asr Ave, Tehran, Iran

Corresponding Author:
Elham Sheykhi Some, MSc
School of Health Management and Information Sciences, Iran University of Medical Sciences, No. 6, Shahid Yasami St, Vali-e-Asr Ave, Tehran, Iran

Abstract

Background: Social media has reshaped health care decision-making; however, its influence on maxillofacial surgeon selection in non-Western contexts such as Iran remains underexplored. Understanding how patients balance digital platforms (eg, Google, Instagram) with traditional referral networks can inform trust dynamics and patient-centered care strategies.

Objective: This study aimed to evaluate the impact of social media compared to personal recommendations on maxillofacial surgeon selection among Iranian patients, assessing decision-making factors, trust perceptions, accuracy concerns, and demographic influences.

Methods: A cross-sectional survey of 384 patients at maxillofacial surgery clinics in Isfahan, Iran (September–November 2023), was conducted using structured questionnaires to collect data on demographics, surgeon selection pathways, social media use, trust, and accuracy concerns. Descriptive statistics, χ^2 tests, one-sample *t* tests, and multiple linear regression were conducted using SPSS Version 26 to analyze platform impact and predictive variables.

Results: Personal recommendations dominated surgeon selection (239/384, 62.2%), significantly outweighing Google (75/384, 19.5%) and Instagram (11/384, 2.9%; χ^2=214.3, *P*<.001). Google and Instagram were used by 160 (41.7%) and 119 (31.0%) patients, respectively; however, their decision-making impact was low with (mean scores: Google 2.27 (0.82), Instagram 2.14 (SD 0.79) on a 1 - 5 scale; *t* tests: *P*<.001). Patient-generated content drove trust, with reviews valued by 144 (37.5%) for Google and 157 (40.9%) for Instagram, and testimonials by 174 (45.3%) for Instagram. Professional credentials influenced 116 (30.2%) participants for Google. Accuracy concerns were moderate; (means values of Google 2.84 (SD 0.91), Instagram 2.85 (SD) 0.88; *P*<.05). Regression identified recommendations (β=.42, *P*<.001), credential trust (β=.19, *P*=.002), and review authenticity (β=.14, *P*=.02) as predictors, while social media use was not a significant predictor (*P*=.32). Participants were predominantly female (233/384, 60.7%), aged 21 - 30 years (117/384, 30.5%), employed (159/384, 41.4%), with moderate income (201/384, 52.3%), and no prior surgery (205/384, 53.4%). Instagram use was higher among younger patients (21 - 30 years: 48/117, 41.0%; χ^2=12.4, *P*=.006).

Conclusions: Social media plays a supplementary role in the selection of maxillofacial surgeons in Iran, with traditional networks prevailing due to cultural trust and low health literacy (adequacy in 43% patients). The emphasis on credible reviews and credentials underscores the need for verified digital content. Contrasting with the digital reliance on aesthetic surgery, these findings advocate for verified profiles, patient education portals, and culturally tailored strategies to enhance trust and patient-centered care.

(JMIR Perioper Med 2025;8:e75899) doi:10.2196/75899

KEYWORDS

social media; patient selection; decision making; trust; oral surgery

Introduction

Social media has transformed health care decision–making by offering patients access to diverse information sources beyond traditional referrals. Platforms such as Google and Instagram play distinct yet complementary roles in shaping provider selection [1].

Google, as a search engine, aggregates patient reviews, professional credentials, and clinic websites, enabling patients to evaluate maxillofacial surgeons based on structured data such as ratings and qualifications [2].

Instagram, a visual platform, showcases before-and-after photos, patient testimonials, and surgeon branding, leveraging aesthetic appeal to influence decisions, particularly in specialties like

maxillofacial surgery, where outcomes are both functional and cosmetic [3].

Globally, 59% of adults use online platforms for health decisions, with 43% relying on patient-generated content [4] . In maxillofacial surgery, where trust and expertise are critical, patients increasingly integrate these digital tools with personal recommendations, although the balance varies by cultural context [3].

The role of Google in health care is pivotal due to its accessibility and comprehensive search capabilities. Approximately 80% of health-related queries begin with Google, with patients seeking surgeon credentials, hospital affiliations, and peer reviews [5]. Google My Business profiles, featuring star ratings and patient feedback, significantly influence trust, with 70% of patients prioritizing high ratings [6]. However, concerns about review authenticity and algorithmic biases persist, as manipulated ratings can mislead patients [7].

Conversely, Instagram conversely, thrives on visual storytelling. In aesthetic surgery, 64% of patients were influenced by Instagram's before-and-after imagery, a trend relevant to maxillofacial surgery [8]. Surgeons use Instagram to post educational content and case studies, achieving high engagement [9]. However, the emphasis of Instagram on aesthetics can oversimplify complex procedures, raising accuracy concerns [10].

In Iran, health care decisions are shaped by collectivist cultural norms, prioritizing familial and community referrals over digital sources. With 53% social media penetration compared to 80% in Western nations, and low health literacy [11], reliance on Google and Instagram is limited.

Iranian patients value verified credentials on Google and authentic testimonials on Instagram, though rural digital access constraints and urban service concentration hinder adoption [12].

Recent studies highlight these dynamics: Zhang et al [13] found that 68% of collectivist society patients prefer offline referrals, while Chegini et al [14] noted trust in peer-endorsed digital content [14].

Despite these insights, a critical research gap remains. No prior studies have quantitatively compared the influence of Google and Instagram versus personal recommendations on maxillofacial surgeon selection in Iran, particularly regarding trust, decision-making factors, and information accuracy. Existing literature focuses on telemedicine or primary care, overlooking specialty-specific dynamics in Iran's culturally distinct health care context [12,15].

This study aims to address this gap by quantitatively assessing the relative influence of Google and Instagram versus personal recommendations on maxillofacial surgeon selection among Iranian patients. By analyzing trust perceptions, decision-making drivers, and accuracy concerns, it seeks to develop evidence-based, culturally tailored strategies to enhance digital trust and patient-centered care in Iran's evolving health care system.

Methods

This descriptive-analytical cross-sectional study was conducted in Isfahan, Iran, a metropolitan city with approximately 2.1 million residents and a center for advanced medical care. A cluster sampling approach was employed for representativeness.

Recruitment took place at three outpatient maxillofacial surgery clinics affiliated with the Faculty of Dentistry at Islamic Azad University, Isfahan, selected for their high patient volume (100 - 150 weekly visits) and diverse patient demographics (socioeconomic and cultural backgrounds). Located in central Isfahan, these clinics offer functional procedures (eg, orthognathic surgery, fracture repair) and cosmetic interventions (eg, facial contouring, genioplasty), serving urban and rural patients.

Data collection took place from September 1 to November 30, 2023, during clinic hours (8:00 AM to 4:00 PM, Sunday to Thursday), using private consultation rooms for confidential survey administration with minimal disruption to operations.

The sample size was calculated based on a population exceeding one million, using a proportion estimate of 0.5, a margin of error of 0.05, and a 95% confidence level, resulting in 384 participants to ensure statistical power and generalizability. No major disruptions (eg, public health restrictions) affected clinic access. Eligible patients were adults aged 18 or older who had visited a maxillofacial surgeon for consultation or surgery and had used social media (eg, Google, Instagram) for health care choices within the past six months, with sufficient literacy to complete the questionnaire. Exclusions included patients under 18, those with cognitive or language barriers (eg, severe developmental disorders, non-Persian speakers without translation support), nonresidents of Iran, those who withdrew consent, provided incomplete responses, or did not use social media for surgeon selection.

Consecutive sampling was applied within selected clinics to reduce bias. Trained receptionists screened patients at check-in using a standardized checklist. Two research assistants approached eligible patients in waiting areas, providing a Persian-language information sheet detailing the study's purpose, voluntary participation, confidentiality, and survey duration (10 - 15 min).

Data Collection Instrument

A structured, researcher-designed questionnaire with 25 items covered four domains:

1. Demographic Information: Age, gender, education, socioeconomic status.
2. Social Media Usage: Frequency of interaction with platforms like Google and Instagram
3. Factors Influencing Surgeon Selection: Impact of peer reviews, social media content, and feedback
4. Satisfaction and Influence: Satisfaction with the chosen surgeon and the role of digital information

Trust and accuracy scores were calculated as mean composite scores from four 5-point Likert-scale items (1=strongly disagree, 5=strongly agree) within the Satisfaction and Influence domain,

assessing the perceived trustworthiness and reliability of social media information used for surgeon selection.

Ten experts confirmed the content's validity (Content Validity Ratio [CVR]=0.8, Content Validity Index [CVI]=1.0). Reliability was assessed via test-retest with 40 patients two weeks apart (Cronbach α=0.704).

Questionnaires were completed in private clinics, with trained assistants ensuring honest responses. Data collection was monitored over three months for consistency and completeness.

Reporting Guideline Adherence

This study adheres to the STROBE (Strengthening the Reporting of Observational studies in Epidemiology) checklist for cross-sectional studies, ensuring transparent reporting of study setting, eligibility criteria, recruitment methods, response rates, and potential biases [16], aligning with Q1 journal reporting standards.

Statistical Analysis

Data were analyzed using descriptive and inferential methods. Demographic variables (eg, age, gender, education) were summarized using means (SD) and percentages. Normality was assessed via Shapiro-Wilk tests. Parametric (eg, two-tailed *t* tests, ANOVA) or nonparametric tests (eg, Mann-Whitney U) were applied based on data distribution. Associations between social media use and surgeon selection were evaluated using Spearman or Pearson correlations. Multiple linear regression models examined the influence of social media usage (eg, Google, Instagram) on decision-making, controlling for age and gender as confounders. Analyses were performed using SPSS software (version 26.0; IBM Corp), with α=.05.

Ethical Considerations

The study protocol underwent ethical review and was approved by the Institutional Review Board (IRB) of Islamic Azad University, Isfahan, under the ethical approval code (IAU.YAZD.REC.1403.092). All 384 participants provided written informed consent before completing the survey.

During recruitment at maxillofacial surgery clinics, trained research assistants provided a Persian-language information sheet outlining the study's purpose, procedures, voluntary participation, and right to withdraw without consequence. Consent forms were signed and collected before survey administration. As this study involved primary data collection, no secondary analysis was performed, rendering additional consent for such purposes inapplicable.

Participant privacy was safeguarded through anonymization of all survey responses at the point of collection. No personally identifiable information (eg, names, addresses, or contact details) was recorded. Surveys were assigned unique numeric codes for data entry. Paper-based surveys were stored in a locked cabinet, and digitized data were maintained on a secure, password-protected server accessible only to authorized researchers. After transcription, paper surveys were securely shredded to prevent data breaches.

No financial or material compensation was offered to participants, given the survey's brief duration (10 - 15 min) and its administration during routine clinic visits. This approach minimized the risk of coercion and aligned with ethical guidelines for low-burden, voluntary research participation.

The manuscript and supplementary materials contain no images of individual participants or users, ensuring no risk of identification. All reported data are aggregated (eg, percentages, means, and statistical summaries), eliminating the need for additional consent for visual content or submission of related consent forms.

Results

This study analyzed responses from 384 patients visiting maxillofacial surgery clinics in Isfahan, Iran, during autumn 2023, to assess the influence of social media on surgeon selection.

Demographic Characteristics and Surgical Experience

The mean age of participants was 29.8 (SD 8.7) years and median of 28 (IQR 18 - 65) years, with a skew toward younger individuals. A higher proportion of participants were female (60.7%). Education was diverse, with most holding a high school diploma or higher (65.1% cumulative). Employment was balanced, with 41.4% employed, and income was moderate, with 45.3% earning 10 - 20 million IRR monthly. Over half (53.4%) had no prior maxillofacial surgery.

Pathways to Surgeon Selection and Social Media Use

Personal recommendations were the primary pathway to surgeon selection, with Google and Instagram playing supplementary roles. A discrepancy in "Pathways to Surgeon" frequencies (374/384 responses) was due to 10 missing responses from an optional question. Social media engagement was moderate, with Google being used more frequently than Instagram in surgeon searches (Table 1).

Table . Frequency and percentage distribution of pathways and social media use for selecting maxillofacial surgeons in a cross-sectional study of 384 patients in Isfahan, Iran, September–November 2023.

Category and subcategories	Participants (N=384) , n (%)
Pathway to surgeon	
Friend or acquaintance recommendation	239 (62.2)
Google Search	75 (19.5)
Instagram	11 (2.9)
Other social media	11 (2.9)
Other	48 (12.5)
Missing responses	
Not applicable	10 (2.6)
Google use for search	
Yes	160 (41.7)
No	224 (58.3)
Instagram use for search	
Yes	119 (31.0)
No	265 (69.0)

Decision-Making and Trust Factors on Social Media

Google users prioritized user reviews and official credentials, while Instagram users valued patient testimonials and real patient experiences (Table 2). χ^2 tests showed significant differences between platforms in decision-making, selection, and trust factors (all P<.001), indicating distinct user behaviors.

Table . Comparison of decision-making, selection, and trust factors influencing maxillofacial surgeon selection on Google and Instagram in a cross-sectional study of 384 patients in Isfahan, Iran, September–November 2023, with χ^2 test results.

Category and subcategories	Participants (Google), (n, %)	Participants (Instagram), (n, %)	χ^2 (*df*)	*P* value
Initial decision-making			117.6 (4)	<.001
User reviews/patient testimonials	135 (35.2)	174 (45.3)		
Surgeon ranking/before-and-after images	85 (22.1)	78 (20.3)		
Website info/Surgeon engagement	32 (8.3)	22 (5.7)		
Photos-videos/educational videos	31 (8.1)	8 (2.1)		
Other	101 (26.3)	102 (26.6)		
Primary selection			76.4 (3)	<.001
Website Info/before-and-after images	30 (7.8)	44 (11.5)		
Rankings/positive reviews	105 (27.3)	157 (40.9)		
Positive reviews/Surgeon interactions	144 (37.5)	56 (14.6)		
Other	105 (27.3)	127 (33.1)		
Trust-building			92.3 (3)	<.001
Authentic Reviews / Patient Experiences	102 (26.6)	153 (39.8)		
Credentials / Procedure Videos	116 (30.2)	78 (20.3)		
Validated Info / Follower Count	53 (13.8)	27 (7.0)		
Other	113 (29.4)	126 (32.8)		

Statistical Analysis of Influence and Concerns

The impact of Google and Instagram on decision-making, measured as mean composite scores from three 1 - 5 Likert items per construct (eg, influence: perceived influence, usefulness, relevance; accuracy: reliability, trust), was low (both *P*s<.001), suggesting limited influence. Moderate concerns about information accuracy were noted (both *Ps*<.05) (Table 3). The large sample size supported parametric tests despite non-normal distributions.

Table . One-Sample *t* test results for impact of information and concerns about accuracy of Google and Instagram in influencing maxillofacial surgeon selection in a cross-sectional study of 384patients in Isfahan, Iran, September–November 2023.

Categories and variables	Participants (N=384), mean (SD)	*t* test (*df*)	*P* value
Impact of information			
Google	2.27 (1.22)	−11.808 (383)	<.001
Instagram	2.14 (1.20)	−14.022 (383)	<.001
Concerns about accuracy			
Google	2.84 (1.18)	−2.628 (383)	.009
Instagram	2.85 (1.26)	−2.315 (383)	.02

Multiple Linear Regression Analysis

Multiple regression analysis (Table 4) revealed that personal recommendations and trust in official credentials significantly predicted decision-making outcomes (*P*<.001and *P*=.002, respectively), while Google and Instagram use did not significantly predict outcomes (*P*>.05). The model fit was acceptable ($F_{5, 378}$=14.32, *P*<.001, R^2=0.16).

Table . Multiple linear regression analysis of factors predicting decision-making outcomes for maxillofacial surgeon selection in a cross-sectional study of 384patients in Isfahan, Iran, September–November 2023, controlling for age and education.

Predictor	β	*t* test	*P* value
Personal recommendations	.42	7.98	<.001
Google use	.09	1.58	.12
Instagram use	.06	1.18	.24
Trust in credentials	.19	3.09	.002
Age	−.03	−0.59	.56
Education	0.04	0.82	.41

Discussion

Principal Results

This study investigated the influence of social media on maxillofacial surgeon selection among 384 patients in Isfahan, Iran, during autumn 2023. The main findings reveal that personal recommendations significantly dominated surgeon selection, with 62.2% of patients relying on friends or acquaintances, compared to 19.5% using Google and only 2.9% using Instagram. While Google (41.7%) and Instagram (31.0%) were used as supplementary tools, their impact on decision-making was significantly below average, as evidenced by mean scores of 2.27 and 2.14, respectively, on a 1 - 5 scale. Patient-generated content, such as reviews and testimonials, alongside professional credentials, emerged as critical factors in decision-making and trust-building across both platforms. However, moderate concerns about information accuracy (means of 2.84 for Google and 2.85 for Instagram, respectively) suggest skepticism toward digital sources, reinforcing the primacy of traditional networks.

Comparison With Prior Work

The prominence of personal recommendations aligns with prior research emphasizing interpersonal trust in health care decisions. The preference for traditional networks in Iran likely stems from cultural factors, such as collectivist values prioritizing familial and community ties, and high social trust in personal referrals over impersonal digital sources. Low health literacy, prevalent in Iran, with only 43% of adults demonstrating adequate health knowledge [17] may further limit reliance on online information, as patients defer to trusted acquaintances for guidance. A recent study by Wang et al [18] found that 68% of patients in collectivist societies preferred offline referrals, mirroring our 62.2% reliance on personal networks, compared to 20% using online platforms. Similarly, Bhatt et al [19] reported that 55% of Asian patients favored word-of-mouth over digital reviews, consistent with our findings.

The supplementary role of social media resonates with global trends. Xiong et al [20] noted that 43% of Chinese patients used online searches; however, offline sources guided final decisions, paralleling our Google (41.7%) and Instagram (31.0%) usage. Zheng et al [21] showed that only 28% of patients globally trust online reviews for surgical choices , aligning with our limited social media impact (P<.001), reinforcing the secondary role of digital platforms. The influence of patient-generated content's (eg, 37.5% for Google reviews, 45.3% for Instagram testimonials) mirrors findings by Thoms et al [22], where 50% of patients valued authentic reviews, though trust was tempered by authenticity concerns, reflected in our moderate accuracy scores (approximately 2.85). This skepticism aligns with Chow et al [23], who reported that 40% of patients questioned digital health information due to a lack of verification.

In contrast, Instagram's minimal role (2.9%) diverges from aesthetic surgery contexts. ElAbd [8] et al found that 64% of plastic surgery patients were influenced by Instagram's visual content, likely due to cosmetic surgery's emphasis on aesthetics unlike the functional focus of maxillofacial procedures in our study. Cultural attitudes in Iran, where social media penetration is lower (53% vs 80% in Western nations) [24], may further reduce Instagram's impact, reinforcing traditional networks.

The role of professional credentials (30.2% for Google) corroborates the findings of Daraz et al [25], where 48% prioritized verified qualifications online. Moderate accuracy concerns highlight a trust gap, consistent with findings by Sorensen et al [26], who noted patients' demand for verified digital content to bridge credibility issues.

Implications for Clinical Practice

The dominance of personal recommendations underscores the need for maxillofacial surgeons to maintain robust referral networks with colleagues and satisfied patients. To enhance their online presence, surgeons should prioritize authentic patient testimonials, verified credentials, and transparent procedure information on platforms like Google and Instagram. For example, creating verified Google Business Profiles with certified reviews or Instagram posts showcasing patient outcomes and qualifications can build trust. Hospitals and clinics could develop patient education portals with validated content to address accuracy concerns, particularly for populations with low health literacy. Professional organizations should advocate for standardized online verification processes, such as digital badges for board-certified surgeons, to counter skepticism. These strategies can complement traditional networks, leveraging digital tools to reach tech-savvy patients while maintaining credibility.

Limitations

The study's focus on Isfahan limits generalizability, as rural or regional variations in Iran may differ. The cross-sectional study design prevents causal inferences about digital influence over time. Additionally, the participants' moderate income and education may not reflect less affluent or less educated groups, potentially underestimating digital access barriers.

Future Directions

Longitudinal studies should track social media's evolving role in Iran as digital literacy improves. Comparative research across surgical specialties and cultural contexts could clarify drivers of platform use. Exploring interventions such as blockchain-verified reviews or artificial intelligence–driven content validation may address accuracy concerns, enhancing digital tools' role in surgeon selection.

Conclusion

In conclusion, personal recommendations remain the cornerstone of maxillofacial surgeon selection in Isfahan, with social media playing a limited, supplementary role shaped by patient reviews and credentials. These findings, consistent with global patterns, yet distinct in their low Instagram reliance, highlight the enduring trust in traditional networks and the need for credible digital enhancements in health care decision-making.

Acknowledgments

We express our sincere gratitude to the 384 patients who participated in this study at maxillofacial surgery clinics in Isfahan, Iran, for their time and willingness to share their experiences. We also thank the clinic staff for their cooperation in facilitating data collection. Special thanks are extended to Grok, developed by xAI, for its invaluable assistance in editing the language of this manuscript, enhancing its clarity and readability. Finally, we appreciate the insightful feedback from colleagues who reviewed early drafts, contributing to the refinement of this work.

All authors declared that they had insufficient funding to support open access publication of this manuscript, including from affiliated organizations or institutions, funding agencies, or other organizations. JMIR Publications provided article processing fee (APF) support for the publication of this article.

Data Availability

The data supporting the findings of this study are not publicly available due to ethical restrictions protecting patient confidentiality. Anonymized data may be provided upon reasonable request to the corresponding author, subject to approval by the institutional ethics committee and compliance with applicable data protection regulations.

Conflicts of Interest

None declared.

References

1. Graham AL, Cobb CO, Cobb NK. The internet, social media, and health decision-making. In: Handbook of Health Decision Science: Springer Science + Business Media, Vol. 2016:335-355. [doi: 10.1007/978-1-4939-3486-7_24]
2. Lali MIU, Mustafa RU, Saleem K, Nawaz MS, Zia T, Shahzad B. Finding healthcare issues with search engine queries and social network data. Int J Semant Web Inf Syst 2017 Jan;13(1):48-62. [doi: 10.4018/IJSWIS.2017010104]
3. Nguyen MN, Yeung JG, Randall Y, Boesze-Battaglia K, Panchal N. Oral and maxillofacial surgery social media boom: potential concerns of social media use for the surgeon. J Oral Maxillofac Surg 2021 Dec;79(12):2396-2397. [doi: 10.1016/j.joms.2021.05.022] [Medline: 34174220]
4. Lim MSC, Molenaar A, Brennan L, Reid M, McCaffrey T. Young adults' use of different social media platforms for health information: insights from web-based conversations. J Med Internet Res 2022 Jan 18;24(1):e23656. [doi: 10.2196/23656] [Medline: 35040796]
5. Kobrinskii BA, editor. Search engine decision-relevant information and exchange with the information system. : Springer Presented at: Proceedings of the Fifth International Scientific Conference "Intelligent Information Technologies for Industry"(IITI'21). [doi: 10.1007/978-3-030-87178-9_31]
6. Chen TT, Wu CI, Chiu MHP, et al. Factors associated with the patient/client use of report cards, physician rating websites, social media, and google for hospital and physician selection: a nationwide survey. Healthcare (Basel) 2022 Oct 1;10(10):1931. [doi: 10.3390/healthcare10101931] [Medline: 36292378]
7. Van Haasteren A. Trust in digital health [Thesis]. : ETH Zurich; 2019.
8. ElAbd R, Alghanim K, Alnesef M, Alyouha S, Samargandi OA. Aesthetic surgery before-and-after photography bias on Instagram. Aesthetic Plast Surg 2023 Oct;47(5):2144-2149. [doi: 10.1007/s00266-023-03398-9] [Medline: 37253847]
9. Ma B, Rojas EM, Li AYZ, Kinard BE. To what extent is oral and maxillofacial surgery educational content posted on Instagram? Int J Oral Maxillofac Surg 2024 Oct;53(10):887-893. [doi: 10.1016/j.ijom.2024.07.013] [Medline: 39079767]
10. Rekawek P, Wu B, Hanna T. Minimally invasive cosmetic procedures, social media, and oral-maxillofacial surgery: use of trends for the modern practice. J Oral Maxillofac Surg 2021 Apr;79(4):739-740. [doi: 10.1016/j.joms.2020.10.038] [Medline: 33259784]
11. Haeri-Mehrizi A, Mohammadi S, Rafifar S, et al. Health literacy and mental health: a national cross-sectional inquiry. Sci Rep 2024 Jun 13;14(1):13639. [doi: 10.1038/s41598-024-64656-7] [Medline: 38871848]

12. Rezaei N, Moghaddam SS, Farzadfar F, Larijani B. Social determinants of health inequity in Iran: a narrative review. J Diabetes Metab Disord 2023 Jun;22(1):5-12. [doi: 10.1007/s40200-022-01141-w] [Medline: 36373156]
13. Zhang S, Zhou H, Zhu Y. Have we found a solution for health misinformation? A ten-year systematic review of health misinformation literature 2013-2022. Int J Med Inform 2024 Aug;188:105478. [doi: 10.1016/j.ijmedinf.2024.105478] [Medline: 38743994]
14. Chegini Z, Kakemam E, Behforoz A, Lotfollah-Zadeh F, Jafari-Koshki T, Khodayari Zarnag R. Impact of patient communication preferences on the patient trust in physicians: a cross-sectional study in Iranian outpatient's clinics. J Patient Exp 2022;9:23743735211069809. [doi: 10.1177/23743735211069809] [Medline: 35024443]
15. Ashtarian K, Etemadi M. Popular diffusion as an instrument for overcoming barriers to digital health in Iran: the critical role of the pandemic. IJHG 2023 Oct 10;28(3):249-266. [doi: 10.1108/IJHG-10-2022-0094]
16. Skrivankova VW, Richmond RC, Woolf BAR, et al. Strengthening the Reporting of Observational Studies in Epidemiology Using Mendelian Randomization: The STROBE-MR Statement. JAMA 2021 Oct 26;326(16):1614-1621. [doi: 10.1001/jama.2021.18236] [Medline: 34698778]
17. Haghdoost AA, Karamouzian M, Jamshidi E, et al. Health literacy among Iranian adults: findings from a nationwide population-based survey in 2015. East Mediterr Health J 2019 Nov 25;25(11):828-836. [doi: 10.26719/emhj.19.017] [Medline: 31782520]
18. Wang R, Huang Y, Zhang X, Yao Y. Online dialogue with medical professionals: an empirical study of an online "Ask the Doctor" platform. Int J Med Inform 2023 Sep;177:105123. [doi: 10.1016/j.ijmedinf.2023.105123] [Medline: 37329764]
19. Bhatt NR, Czarniecki SW, Borgmann H, et al. A systematic review of the use of social media for dissemination of clinical practice guidelines. Eur Urol Focus 2021 Sep;7(5):1195-1204. [doi: 10.1016/j.euf.2020.10.008] [Medline: 33172773]
20. Xiong Z, Zhang L, Li Z, Xu W, Zhang Y, Ye T. Frequency of online health information seeking and types of information sought among the general Chinese population: cross-sectional study. J Med Internet Res 2021 Dec 2;23(12):e30855. [doi: 10.2196/30855] [Medline: 34860676]
21. Zheng B, Cai X. Fundamentals of digital surgery: surgeon-centered data enchantment for presurgical planning, intraoperative performance and decision making. Laparosc Endosc Robot Surg 2025. [doi: 10.1016/j.lers.2025.02.005]
22. Thoms B, Botts N, Eryilmaz E. Examining trust and consent models for patient-generated health data-sharing and incentives. Presented at: Hawaii International Conference on System Sciences; Jan 7-10, 2025. [doi: 10.24251/HICSS.2025.418]
23. Chow E, Virani A, Pinkney S, et al. Caregiver and youth characteristics that influence trust in digital health platforms in pediatric care: mixed methods study. J Med Internet Res 2024 Oct 28;26:e53657. [doi: 10.2196/53657] [Medline: 39467279]
24. Ahamed A, Gong W. What drives high penetration rates of social media a qualitative comparative analysis across countries. JBM 2022;28(1):101-130. [doi: 10.1504/JBM.2022.141296]
25. Daraz L, Morrow AS, Ponce OJ, et al. Can Patients trust online health information? A meta-narrative systematic review addressing the quality of health information on the internet. J Gen Intern Med 2019 Sep;34(9):1884-1891. [doi: 10.1007/s11606-019-05109-0] [Medline: 31228051]
26. Sørensen K. Fostering digital health literacy to enhance trust and improve health outcomes. Computer Methods and Programs in Biomedicine Update 2024;5:100140. [doi: 10.1016/j.cmpbup.2024.100140]

Edited by N Rohatgi; submitted 12.04.25; peer-reviewed by F Yeh, S Na; revised version received 06.06.25; accepted 10.06.25; published 14.08.25.

Please cite as:
Abrishami M, Bayat M, Sheykhi Some E
Social Media Influence on Surgeon Selection Among Iranian Maxillofacial Patients: Cross-Sectional Survey Study
JMIR Perioper Med 2025;8:e75899
URL: https://periop.jmir.org/2025/1/e75899
doi:10.2196/75899

Barriers to Wellness Among General Surgery Residents During the COVID-19 Pandemic: Qualitative Analysis of Survey Responses

Idil Bilgen[1*], MD; Matthew Castelo[2,3*], MD, PhD; Emma Reel[2], MSW; May-Anh Nguyen[2], MD; Brittany Greene[2], MD; Justin Lu[4], BHSc; Savtaj Brar[2], MD; Tulin Cil[2,5*], MD

[1]School of Medicine, Koç University, Istanbul, Turkey

[2]Department of Surgery, University of Toronto, Toronto, ON, Canada

[3]Institute of Health Policy, Management and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

[4]Temerty Faculty of Medicine, University of Toronto, Toronto, ON, Canada

[5]Princess Margaret Cancer Centre, University Health Network, OPG Wing, 6th Floor, 5W278, 610 University Ave, Toronto, ON, Canada

[*]these authors contributed equally

Corresponding Author:
Tulin Cil, MD
Department of Surgery, University of Toronto, Toronto, ON, Canada

Abstract

Background: Health care provider burnout worsened during the COVID-19 pandemic.

Objective: This qualitative study described general surgery residents' perceptions of burnout and the impact of the COVID-19 pandemic and their attitudes toward wellness initiatives.

Methods: General surgery residents at a large training program in Canada completed a 21-item survey focused on self-reported burnout, mental health, perceptions of wellness resources, and the effects of the COVID-19 pandemic. Free-text responses were extracted for qualitative thematic content analysis. A coding framework was established, and emergent themes were identified.

Results: A total of 62% (51/82) of the residents completed the survey. Most respondents were senior residents (21/51, 41%) and identified as male (32/51, 63%). In total, 65% (33/51) of the residents met the criteria for burnout. Three themes were identified: (1) the culture of general surgery does not promote wellness, (2) the COVID-19 pandemic worsened existing access to vacation days and rest, and (3) wellness education in general surgery is ineffective and onerous to complete. General surgery residents emphasized the rigid lifestyle and culture of the specialty. Residents said that the idea of wellness was poorly executed. COVID-19 protocols increased the acceptance of taking sick days, but this was offset by staff shortages during the pandemic. Finally, residents emphasized the inefficacy of wellness education. They felt that they did not lack knowledge on reaching wellness but simply lacked the adequate time and resources to improve their well-being.

Conclusions: There are persistent concerns within the culture of general surgery that were further impacted by workload and stress during the pandemic. These results may inform future programmatic efforts to decrease resident burnout.

(JMIR Perioper Med 2025;8:e72819) doi:10.2196/72819

KEYWORDS

burnout; COVID-19; general surgery; internship and residency; qualitative

Introduction

Burnout is an occupational syndrome characterized by emotional exhaustion, detachment, poor self-worth, and low personal accomplishment [1]. It is common among medical professionals [2,3] and has been associated with worse patient care, increased medical errors, poor physician mental health, and training attrition [4-7]. A particularly vulnerable group of physicians is surgical trainees, as identified in several high-impact publications [2,8,9]. General surgery residents undergo an extended period of training (5 or more years) and are subjected to long work hours and chronic sleep deprivation. Indeed, these are among the reasons most often cited by trainees seriously considering leaving general surgery residency [10]. However, efforts to restrict duty hours have not clearly demonstrated benefit [11], and it has become clear that burnout in general surgery training is a complex phenomenon. Other factors such as discrimination and sexual harassment experienced during training may certainly contribute to both attrition and burnout [8].

The COVID-19 pandemic further exacerbated many of these preexisting issues while introducing new stressors to the training

environment [12,13]. General surgery residents were called upon to assist in overwhelmed intensive care units (ICUs) during successive COVID-19 waves. Aside from missed surgical experience and disruption of surgical rotations, the stress of working with patients with COVID-19 was a contributing factor to physician burnout [14-18]. The negative effects of the pandemic have been reflected in the physical health, safety, and mental well-being of residents [19-25]. While there are emerging data on the impact of COVID-19 on resident wellness [12,13], there are minimal data on understanding residents' experiences of the pandemic and their relationship with the culture of general surgery training more broadly. Given the larger systemic issues within training programs beyond work hours and redeployment, it is important to generate resident-specific data [26].

The aim of this study was to understand general surgery residents' perceptions of burnout and the interaction with the COVID-19 pandemic experience through qualitative analysis of a wellness survey conducted at a large university-based training program.

Methods

Study Design

This was a qualitative descriptive study [27,28] using free-text responses to a cross-sectional survey conducted within the General Surgery residency training program at the University of Toronto.

Participants

The General Surgery residency program at the University of Toronto is the largest such training program in Canada, comprising 84 clinical and research residents completing full-time graduate degrees [29]. The survey was administered through SurveyMonkey (SurveyMonkey Inc) to all registered general surgery residents as of February 16, 2021. Two additional email reminders were sent 2 weeks apart. Two residents who developed the survey were excluded (MAN and BG).

Study Team

The study team primarily responsible for designing the survey included SB (the program director of the General Surgery residency program at the University of Toronto), TC (the wellness lead and active researcher in surgical education), MAN (general surgery senior resident), and BG (general surgery senior resident). IB and ER conducted the qualitative analysis and thematic coding. ER is an experienced social worker with expertise in qualitative methodology. MC is a surgeon and clinical epidemiologist who conducted the statistical analysis.

Survey Development and Content

The original objectives of the survey were to assess resident burnout, self-reported mental health, and perceptions of available wellness and support resources in response to the COVID-19 pandemic. These results have been reported separately [30]. At the time of survey administration, Ontario was experiencing ongoing COVID-19–related disruption with widespread public health restrictions. The period immediately before the survey (December 2020 and January 2021) was marked by a wave of SARS-CoV-2 infections, high ICU use, and residency training interruption and redeployment [31]. This directly led to the development of this survey.

The 21 survey questions were developed and reviewed for clarity and sensibility. The entire survey is shown in Multimedia Appendix 1. It consisted of questions reporting demographics, level of training, perceived importance of wellness, the efficacy of wellness initiatives undertaken by the residency program, the availability of such resources, and regret for pursuing surgical residency. We also measured self-identified burnout, formal burnout measures, and mental health concerns. The survey was transferred to an online format [32] and pilot-tested before administration. A total of 81 free-text responses were available for 17 survey questions, which underwent coding and formed the data source for this qualitative analysis.

Burnout was formally assessed using the Maslach Burnout Inventory–Human Services Survey for Medical Personnel (MBI-HSS) [1]. The MBI-HSS is a validated psychometric tool created to study burnout in medical professionals and assesses 3 dimensions of burnout: emotional exhaustion, depersonalization, and personal accomplishment (Multimedia Appendix 1). Higher scores for emotional exhaustion and depersonalization, along with lower scores for personal accomplishment, indicate higher degrees of burnout. The presence of burnout according to the MBI-HSS was defined in accordance with multiple high-impact studies [3,8,33,34]. Participants were considered to have burnout if they had at least weekly symptoms on a subset of 6 scale items that form the abbreviated MBI-HSS [35].

Analysis

Participant characteristics were tabulated and analyzed using descriptive statistics. The survey questions and corresponding free-text responses were extracted for qualitative thematic content analysis [36]. Comments were coded into specific themes. An inductive and semantic qualitative approach was undertaken to obtain a thorough grasp of the residents' perceptions of wellness and burnout in general surgery during the pandemic [27,28,37]. A reflexivity practice was embedded in the analysis process. Analysts engaged in memoing while independently reviewing the data; these notes were discussed alongside the data as part of peer debriefing, during which themes and analyses were discussed within the study team and a coding framework was established based on consensus [27,28,36]. Furthermore, the analysts reflected on their positions within the research and how their own educational and professional identities may impact their interpretation of the data [38]. A primary analyst coded all the transcripts, with a second analyst coding a subset independently using Microsoft Excel (IB and ER).

Ethical Considerations

The Research Ethics Board at the University of Toronto reviewed this study and granted ethics approval (study #40135). Informed consent was obtained from all participants, and they were given an opportunity to opt out at any time. Their participation in this study was not mandatory. The data were deidentified, and no compensation was provided.

Results

Participant Demographics and MBI-HSS

The survey was distributed to 82 general surgery residents at the University of Toronto, and 51 (62%) completed it. Demographics are presented in Table 1. Most respondents were senior residents (21/51, 41%), followed by junior residents (18/51, 35%) and research residents (11/51, 22%). Most respondents identified as male (32/51, 63%). Of note, the gender breakdown of the program at the time of the survey was 60% male. There was a nonsignificant increase in residents identifying themselves as having burnout during the pandemic compared to before the pandemic (32/51, 63% vs 29/51, 57%, respectively; *P*=.21). Most respondents (29/51, 57%) never or rarely regretted training in general surgery, whereas 37% (19/51) sometimes regretted it and only 6% (3/51) usually regretted it.

Table . Participant demographics (N=51).

Characteristic	Participants, n (%)
Year of training	
PGY[a] 1	6 (12)
PGY 2	12 (24)
PGY 3	9 (18)
PGY 4	5 (10)
PGY 5	7 (14)
Research	11 (22)
Preferred not to answer	1 (2)
Level of training	
Junior resident	18 (35)
Senior resident	21 (41)
Research	11 (22)
Preferred not to answer	1 (2)
Self-identified gender	
Female	19 (37)
Male	32 (63)
Self-described burnout before the pandemic	
Maybe	7 (14)
No	15 (29)
Yes	29 (57)
Self-described burnout during the pandemic	
Maybe	10 (20)
No	9 (18)
Yes	32 (63)
Regret for general surgery residency	
Never	15 (29)
Rarely	14 (27)
Sometimes	19 (37)
Usually	3 (6)

[a]PGY: postgraduate year.

MBI-HSS subscale scores are presented in Table 2. They indicated that personal achievement occurred between once per week and a few times per week (median 4.62/7, IQR 4.00-5.31), emotional exhaustion occurred between once per month and a few times per month (median 2.67/7, IQR 1.94-3.44), and depersonalization occurred between a few times per year and once per month (median 1.80/7, IQR 1.20-2.50). A total of 65% (33/51) of the residents met the criteria for burnout using the MBI-HSS.

Table . Participant burnout scores (N=51).

	Values
Met criteria for burnout, n (%)	33 (65)
MBI-HSS[a]—emotional exhaustion score (range 0-6), median (IQR)	2.67 (1.94-3.44)
MBI-HSS—personal achievement score (range 0-6), median (IQR)	4.62 (4.00-5.31)
MBI-HSS—depersonalization score (range 0-6), median (IQR)	1.80 (1.20-2.50)

[a]MBI-HSS: Maslach Burnout Inventory–Human Services Survey for Medical Personnel.

Emergent Themes

Overview

Our analysis identified 3 emergent themes in the free-text survey responses (Table 3). These were (1) the culture of general surgery does not promote wellness, (2) the COVID-19 pandemic worsened access to time off, and (3) wellness education in general surgery is ineffective and onerous to complete. These themes are presented with supporting quotes in the following sections.

Table . Emergent themes and selected quotes.

Theme	Selected quotes
The culture of general surgery does not promote wellness	• "I am not aware of wellness resources available specifically from the General Surgery department." [Participant ID 35] • "Residents were left out of the picture on many occasions. Mass emails to the resident body not very helpful." [Participant ID 25] • "Wellness depends on having free time. Ie time to speak to a psychologist, see a doctor, rest, engage with family etc. Wellness curriculum/didactic teaching is further inhibitory to resident access to the above activities (especially much needed appointments) by further scheduling activities into already full schedules." [Participant ID 7]
The COVID-19 pandemic worsened existing access to vacation days and rest	• "Pre-COVID times—came in while still sick out of perceived expectation." [Participant ID 34] • "...never even bothered to try—rumors about how rarely they were approved, plus seeing people already doing at—or over—PARO max call." [Participant ID 19] • "I was discouraged from taking lieu days early in my postgraduate training." [Participant ID 23]
Wellness education in general surgery is ineffective and onerous to complete	• "...any format with flexibility—forced wellness activities feel counter-intuitive." [Participant ID 19] • "Give residents protected time away from hospital/lecture hall in casual/relaxed setting. Don't educate re: wellness; facilitate wellness." [Participant ID 22] • "As residents, we have little to no control over our schedules, or our life in general. Wellness education and initiatives need to come from the top down and be incorporated within the residency program. For example, taking a lieu day should not be something we necessarily ask for, however, when a chief resident schedules a resident to be on call on a holiday, they should subsequently ask when they want to take their lieu day. This type of approach would ensure that the resident feels comfortable taking the lieu day, and remove any sense of guilt they may feel taking a lieu day. When it comes to wellness education, I think the goals are two-fold. 1. Education at the higher levels so that individuals in positions of power, eg, Staff, Fellows, Chief Residents, Sr. Residents, are educated on how to create an environment and workplace that promotes wellness for everyone involved. 2. Individual level education on how we can each individually engage in activities to promote our wellness." [Participant ID 32] • "Increasing the number of the residents in the program might help decrease the stress and load of work on the residents and number of calls." [Participant ID 43]

The Culture of General Surgery Does Not Promote Wellness

General surgery residents emphasized the rigid lifestyle and culture of their chosen specialty. Multiple residents said that the idea of wellness existed only theoretically and was not well executed in reality. Residents admitted that achieving wellness appeared unlikely due to the perceived expectations of general surgery residency, claiming that "true wellness is somewhat contradictory to the classic culture of surgical residency" (participant 23):

> *I'm not sure how much wellness initiatives targeted at residents specifically will help promote their wellness while the culture of surgery and residency remains what it is. This is a very difficult and complex issue to address.* [Participant 23]

In terms of suggested changes, residents reported that feeling supported by hospital coworkers (eg, staff surgeons, allied health professionals, and administrators) would provide them with a greater sense of security. They felt that achieving wellness requires respect from hospital staff, improvements in the work environment, and a deeper appreciation of efforts and personal sacrifices:

> *Wellness comes from perceiving high self-worth and that our work environment finds us valuable. If I'm treated with respect and valued at work that will generate resilience against burn out... Wellness comes from the program/work environment address[ing] that concern and changing for the better. No podcast/retreat will ever address those concerns adequately.* [Participant 8]

One aspect of surgical culture that was raised was the lack of time to focus on wellness; the perception was that didactic or educational interventions (eg, mandatory retreats) were at times contradictory to personal efforts at wellness. Residents acknowledged that greater recognition and valuation of this need could improve wellness:

> *No more lectures, no more retreats. No more anything that already takes away from our precious little spare time to be human. Encourage and foster and promote more spare time away from work/study to be human and do things that let us recharge. ...If resident wellness programs actually wanted to promote wellness, they would target the systematic issues that makes residency sometimes needlessly grueling as opposed to further taking away our time.* [Participant 16]

Despite these challenges, one resident did feel that it was "reassuring that the General Surgery program has a Wellness Lead and is working towards improving wellness..." [participant 32]. Specific mentors were highlighted as being supportive, particularly the program directors.

In particular, staff mentors were identified as being important resources during the pandemic, with specific staff surgeons being named. However, one resident noted that they were "left out of the picture on many occasions" (participant 25). This may suggest that the program's response to COVID-19 worsened resident involvement and dissemination of information.

The COVID-19 Pandemic Worsened Existing Access to Time Off

Residents highlighted challenges in access to vacation and lieu days during the COVID-19 pandemic. As discussed previously, they emphasized the lack of adequate time for self-care due to long work hours and felt at times undervalued. Interestingly, the pandemic increased the perceived acceptance of taking sick days, likely due to strict COVID-19 protocols implemented at hospitals and training sites. Before these changes, participants reported that it was rare to use sick time:

> *...pre-COVID times—came in while still sick out of perceived expectation.* [Participant 19]

> *I had never taken a sick day until during COVID, when I had a fever for a few days and had to wait for my test result to come back. Before then, I have certainly worked while sick and wished I was at home-but never asked to leave/stay home. Unsure what the response would have been if I had asked to go home or not to come in because I wasn't feeling well. It actually felt nice during COVID to be forced to stay home until my test result came back. I got to rest and recover instead of just powering through...* [Participant 33]

However, this was offset by cancelled vacation days as a result of COVID-19 health care personnel shortages. While residents are entitled to an annual 4 weeks of vacation time, some participants noted that they "only took 2 weeks of vacation due to COVID scheduling" (participant 18):

> *Lieu days quite difficult to take due to clinical demands... Have never taken an educational day in residency. Took 2/4 weeks' vacation (rest was cancelled due to pandemic).* [Participant 40]

In addition to sick days and vacation days, lieu days were also discussed. As per the employment contracts relevant to the study population, residents are eligible for a paid day off if required to work on a recognized holiday to be taken within 90 days of the holiday [39]. Participants admitted that they felt that lieu days only existed theoretically—there was a perceived expectation not to use the paid day off as doing so would worsen existing coverage shortages and would be unfair to coresidents. Some went as far as saying that they were "discouraged to take [lieu days]" (participant 44) and that "it is not acceptable in my residency culture" (participant 41):

> *I have very rarely seen anyone in [the] general surgery program (other than off service) actually use a lieu day. Given how short residents we are at most sites, taking lieu days feels like it will penalize the other residents, who will have to cover.* [Participant 35]

Suggestions offered by participants included streamlining requests for vacation time so that they were centralized and coordinated, encouraging residents to plan lieu days when scheduled to work holidays, and avoiding ward rounds before scheduled teaching activities:

> *When a chief resident schedules a resident to be on call on a holiday, they should subsequently ask when they want to take their lieu day.* [Participant 32]

Wellness Education in General Surgery Is Ineffective and Onerous to Complete

Residents discussed the inefficacy of wellness education. There were differences in how residents perceived wellness and how it was conveyed by the residency program. While many initiatives focused on wellness education or skills, study participants emphasized that they preferred more free time rather than "wellness through mandatory modules and activities" (participant 35):

> *Promoting wellness starts with valuing residents' time. Improving service to education ratio would allow us to focus more on education, and it would reduce our overall hours spent in hospital, which in turn could open up the opportunity for wellness related activities.* [Participant 35]

Participants were not enthusiastic about wellness programs and, instead, preferred time for self-care. It was noted that scheduling personal health appointments (including counseling) was challenging, and many residents completed these activities on postcall days:

> *Wellness depends on having free time, ie, time to speak to a psychologist, see a doctor, rest, engage with family etc. Wellness curriculum/ didactic teaching is further inhibitory to resident access to the above activities (especially much needed appointments) by further scheduling activities into already full schedules.* [Participant 7]

On the basis of their free-text responses, residents felt that they did not lack knowledge on reaching wellness but simply lacked the adequate time and resources for carrying out wellness-promoting behaviors. For this reason, many were against mandatory or scheduled educational sessions as they did not attempt to address the "root causes of unwellness" (participant 24):

> *The ultimate source of burnout is the work burden and not the lack of knowledge on how to be well. We require time to be well. It would be useful if we had academic half days that are dedicated for resident wellness (time for us to make and attend appointments when we are not post call, counseling sessions, catch up on life in general).* [Participant 18]

However, not all wellness-related activities were considered ineffective. Social events held by the program that encouraged socializing with coresidents were praised, although the challenges posed by COVID-19 regarding these types of events were acknowledged. One resident encouraged virtual events instead:

> *Program can facilitate a virtual social like a cooking class or trivia or other ways to engage residents on a program level.* [Participant 51]

Rather than wellness education aimed at residents, some suggested training that could "focus on staff understanding [of] resident wellness" (participant 44). Finally, some participants suggested that wellness activities could take the form of more concrete actions toward clinical stressors such as unnecessary consultations and pages and "respect[ing] academic half days (no rounding before teaching)" (participant 39). One resident noted that call requirements could be lightened to improve wellness:

> *Doing one weekend call/month would be ideal for wellness.* [Participant 43]

Discussion

This cross-sectional qualitative study of free-text responses among general surgery residents surveyed about burnout and wellness initiatives during the COVID-19 pandemic revealed substantial concerns regarding surgical culture, availability of free time, and the perception of wellness initiatives. Residents identified interconnected issues that resulted in little free time to focus on their well-being, and these problems were exacerbated by mandatory modules and educational sessions that were not perceived as helpful and further impeded time for rest.

Surgery has long been recognized as having a rigid culture with heavy workload, long hours, high call demands, and complex patients [40]. There has been increasing recognition of these issues, but solutions are unclear. A major focus of research and policy has been on duty hour restrictions, which have been widely used in the United States and Europe but less so in Canada. These initiatives were spurred by concerns about patient safety and resident wellness over 20 years ago [41]. With time, evidence has accumulated suggesting that duty hour restrictions do not have a strong impact on resident wellness in surgery [11,42]. Instead, as the results of our study indicate, there are broader problems within surgical culture that contribute to burnout. In 2018, Hu et al [8] conducted a cross-sectional survey of the American general surgery resident population (7409 trainees), with 32% reporting gender-based discrimination, 17% reporting racial-based discrimination, and 30% reporting verbal or physical abuse during training. While our participants did not raise concerns about mistreatment or abuse directly, they pointed to systemic issues such as feeling undervalued and unsupported. Evidence has pointed to deep-rooted cultural issues within the surgical environment that contribute to bullying and burnout. These include hierarchical power structures and problematic individual surgeons with narcissistic tendencies that support these dynamics [43,44]. In a comprehensive examination of this issue, Munro and Phillips [45] quote evidence that surgical trainees are 3 times more likely to experience bullying compared to other trainees. Reasons again include the essential power imbalance that comes from a hierarchical work environment, the public nature of feedback in the operating room, and emulation of bullying behavior itself as a consequence of prolonged training years.

Our study demonstrated that, during the COVID-19 pandemic, some of the barriers to wellness were exacerbated. Although the survey was conducted during the pandemic and residents had experienced at least 2 substantial waves with increased pressures on hospitals and redeployments, participants did not

directly relate these events to burnout. Rather, they raised concerns surrounding the cancellation of vacation time and identified issues with receiving entitled lieu days. Literature from early in the pandemic has shown a substantial impact on residents' educational opportunities and well-being [12-14]. It is possible that our participants were affected by recall bias or they intentionally chose to raise broader issues with respect to burnout rather than focusing on the pandemic specifically. Furthermore, junior residents would have had limited to no prepandemic experience with residency training, thereby not having a comparative experience outside of the COVID-19 pandemic years. Finally, some stressors that are cited in this early pandemic literature may have faded, including poor access to personal protective equipment and fear of infection. In addition to vaccination providing robust protection from serious complications of COVID-19, our resident population was relatively young and healthy and could be expected to experience mild illness. This is an important finding as programs may need to shift away from wellness initiatives directed at pandemic stressors back to preexisting systemic issues in surgery.

There is literature showing that both individual and structural wellness interventions can reduce burnout among physicians [26]. Most of these take the form of mindfulness or stress reduction sessions and small group discussions. Our results suggest that these interventions would be poorly perceived by surgical residents, who felt that these further placed time demands on busy residents and may worsen burnout. Randomized studies of structural and organizational interventions are less common [11,46-48], but these were more widely supported by our participants and may have more success in this population. In a landmark study published in *The New England Journal of Medicine*, Bilimoria et al [11] randomized 117 general surgery residency programs in the United States to follow strict duty hour restrictions or a more flexible policy that waived rules on shift length and time between shifts. Residents working under the flexible policy were less likely to report patient safety concerns or negative effects on professionalism and education but noted a greater impact on their personal activities. Overall, the trial questioned the evidence base behind restrictive duty hour policies. In a similar trial conducted in Canada, ICU residents were randomized to overnight shifts between 12 and 24 hours in length [48]. Somatic symptoms among residents were more common in the 24-hour shift group, but rates of burnout were similar. There was a tendency toward more preventable errors in patient care occurring in the 12-hour shift group. This literature highlights tension between patient safety and resident wellness and, similar to our results, suggests that institutional policy may be less effective than broader cultural change and sense of belonging at improving burnout [49]. Further research as to the optimal ways to address burnout among trainees is critically needed.

This study has several strengths. Our response rate was high, indicating engagement with the residency program. We used formal qualitative analysis to explore the free-text responses to our survey rather than simple quantitative methods [36,50]. Our assessment of burnout was complemented by formal measures of burnout (MBI-HSS scores), and we have previously described these results [30]. Although data resulting from in-depth interviews may have been richer, given the limitations of conducting research during the pandemic, we were still able to secure a detailed set of responses from a relatively large group of general surgery residents who had experienced considerable disruption in training due to COVID-19. Our study is an important contribution to our understanding of barriers to improving wellness among general surgery residents.

This study has limitations. While the survey was distributed to 82 residents in the program, 51 (62%) completed it, and not all participants chose to offer free-text responses. Thus, our study cannot be considered a random sample of residents in the program. Those with particularly negative or strong opinions or experiences may have been more likely to use the free-text responses and have their views represented in this study. However, we found remarkable consistency in the responses, and the issues raised were not unexpected. As discussed previously, survey responses do not constitute as detailed a data source as interviews [51], although qualitative analyses of surveys are common. This approach is supported by the literature and the methodology of other similar studies, with free-text answers shown to be a rich source of data [52-55]. Inductive analysis is data-driven and an efficient method [37]. Our methodology maps closely to the recommendations of Thomas [56], including a close reading of text, creation of categories and themes, overlapping coding and revisiting uncoded text, and continuing revision and refinement of the category system. We also conducted consistency checks, such as independent parallel coding, checking on the clarity of categories, and research team discussions [56]. However, we still plan to conduct in-depth telephone interviews to further explore the findings generated by this study. We did not provide participants with formal definitions for terms such as "wellness" or "wellness education." Thus, there may have been heterogeneity in how participants understood these core concepts. These data, particularly with respect to COVID-19, may not be generalizable to other programs. At the time the survey was completed, Toronto had greater experience with COVID-19 in health care settings than other Canadian cities, and the residents likely had more disruption to their training compared to colleagues in less dense centers. Conversely, they were likely less disrupted compared to American programs in similarly large cities—Canada had far less COVID-19–related morbidity on a per capita basis compared to the United States [57]. The residency program at the University of Toronto is also highly academic, with a larger number of research residents and unique work environments that may further limit generalizability to other institutions. A final limitation of our analysis is that we did not conduct member checking, where the conclusions of the study are fed back to the respondents to ensure validity [58].

This qualitative analysis of survey responses among general surgery residents working during the COVID-19 pandemic led to the recognition of 3 important themes related to wellness: the friction between the culture of surgery and wellness prioritization, the impact of the pandemic on working hours, and the difficulties with wellness interventions. These results represent the resident experience at a unique time in health care

and may inform future efforts to design interventions to decrease burnout.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
General surgery residency wellness and burnout survey.
[DOCX File, 85 KB - periop_v8i1e72819_app1.docx]

References

1. Maslach C, Jackson SE, Leiter MP. The Maslach Burnout Inventory Manual: Consulting Psychologists Press; 1996.
2. Dyrbye LN, West CP, Satele D, et al. Burnout among U.S. medical students, residents, and early career physicians relative to the general U.S. population. Acad Med 2014 Mar;89(3):443-451. [doi: 10.1097/ACM.0000000000000134] [Medline: 24448053]
3. West CP, Shanafelt TD, Kolars JC. Quality of life, burnout, educational debt, and medical knowledge among internal medicine residents. JAMA 2011 Sep 7;306(9):952-960. [doi: 10.1001/jama.2011.1247] [Medline: 21900135]
4. Shanafelt TD, Noseworthy JH. Executive leadership and physician well-being: nine organizational strategies to promote engagement and reduce burnout. Mayo Clin Proc 2017 Jan;92(1):129-146. [doi: 10.1016/j.mayocp.2016.10.004] [Medline: 27871627]
5. Shanafelt T, Goh J, Sinsky C. The business case for investing in physician well-being. JAMA Intern Med 2017 Dec 1;177(12):1826-1832. [doi: 10.1001/jamainternmed.2017.4340] [Medline: 28973070]
6. Fahrenkopf AM, Sectish TC, Barger LK, et al. Rates of medication errors among depressed and burnt out residents: prospective cohort study. BMJ 2008 Mar 1;336(7642):488-491. [doi: 10.1136/bmj.39469.763218.BE] [Medline: 18258931]
7. Dyrbye LN, Massie FS Jr, Eacker A, et al. Relationship between burnout and professional conduct and attitudes among US medical students. JAMA 2010 Sep 15;304(11):1173-1180. [doi: 10.1001/jama.2010.1318] [Medline: 20841530]
8. Hu YY, Ellis RJ, Hewitt DB, et al. Discrimination, abuse, harassment, and burnout in surgical residency training. N Engl J Med 2019 Oct 31;381(18):1741-1752. [doi: 10.1056/NEJMsa1903759] [Medline: 31657887]
9. Khorfan R, Hu YY, Agarwal G, et al. The role of personal accomplishment in general surgery resident well-being. Ann Surg 2021 Jul 1;274(1):12-17. [doi: 10.1097/SLA.0000000000004768] [Medline: 33491973]
10. Gifford E, Galante J, Kaji AH, et al. Factors associated with general surgery residents' desire to leave residency programs: a multi-institutional study. JAMA Surg 2014 Sep;149(9):948-953. [doi: 10.1001/jamasurg.2014.935] [Medline: 25075473]
11. Bilimoria KY, Chung JW, Hedges LV, et al. National cluster-randomized trial of duty-hour flexibility in surgical training. N Engl J Med 2016 Feb 25;374(8):713-727. [doi: 10.1056/NEJMoa1515724] [Medline: 26836220]
12. Poelmann FB, Koëter T, Steinkamp PJ, Vriens MR, Verhoeven B, Kruijff S. The immediate impact of the coronavirus disease 2019 (COVID-19) pandemic on burn-out, work-engagement, and surgical training in the Netherlands. Surgery 2021 Sep;170(3):719-726. [doi: 10.1016/j.surg.2021.02.061] [Medline: 33820653]
13. Coleman JR, Abdelsattar JM, Glocker RJ, RAS-ACS COVID-19 Task Force. COVID-19 pandemic and the lived experience of surgical residents, fellows, and early-career surgeons in the American College of Surgeons. J Am Coll Surg 2021 Feb;232(2):119-135.e20. [doi: 10.1016/j.jamcollsurg.2020.09.026] [Medline: 33069850]
14. Aziz H, James T, Remulla D, et al. Effect of COVID-19 on surgical training across the United States: a national survey of general surgery residents. J Surg Educ 2021;78(2):431-439. [doi: 10.1016/j.jsurg.2020.07.037] [Medline: 32798154]
15. Ammann AM, Cortez AR, Vaysburg DM, et al. Examining the impact of COVID-19 restrictions on the operative volumes of US general surgery residents. Surgery 2022 Feb;171(2):354-359. [doi: 10.1016/j.surg.2021.06.003] [Medline: 34247838]
16. Clements JM, Burke JR, Hope C, et al. The quantitative impact of COVID-19 on surgical training in the United Kingdom. BJS Open 2021 May 7;5(3):zrab051. [doi: 10.1093/bjsopen/zrab051] [Medline: 34169311]
17. Coons BE, Tam SF, Okochi S. Rapid development of resident-led procedural response teams to support patient care during the coronavirus disease 2019 epidemic: a surgical workforce activation team. JAMA Surg 2020 Aug 1;155(8):683-684. [doi: 10.1001/jamasurg.2020.1782] [Medline: 32352481]
18. Nassar AH, Zern NK, McIntyre LK, et al. Emergency restructuring of a general surgery residency program during the coronavirus disease 2019 pandemic: the University of Washington experience. JAMA Surg 2020 Jul 1;155(7):624-627. [doi: 10.1001/jamasurg.2020.1219] [Medline: 32250417]

19. Kannampallil TG, Goss CW, Evanoff BA, Strickland JR, McAlister RP, Duncan J. Exposure to COVID-19 patients increases physician trainee stress and burnout. PLoS ONE 2020 Aug 6;15(8):e0237301. [doi: 10.1371/journal.pone.0237301] [Medline: 32760131]
20. Holzer KJ, Lou SS, Goss CW, et al. Impact of changes in EHR use during COVID-19 on physician trainee mental health. Appl Clin Inform 2021 May;12(3):507-517. [doi: 10.1055/s-0041-1731000] [Medline: 34077972]
21. Huff A, Niedermier J, Kelm Z, Jara A, Barnett L. Dynamic trainee support for residents involved in COVID-19 treatment and response. Psychiatr Serv 2020 Jul 1;71(7):753. [doi: 10.1176/appi.ps.71703] [Medline: 32605506]
22. Giacalone A, Rocco G, Ruberti E. Physical health and psychosocial considerations during the coronavirus disease 2019 outbreak. Psychosomatics 2020;61(6):851-852. [doi: 10.1016/j.psym.2020.07.005] [Medline: 32861431]
23. Tracy DK, Tarn M, Eldridge R, Cooke J, Calder JD, Greenberg N. What should be done to support the mental health of healthcare staff treating COVID-19 patients? Br J Psychiatry 2020 Oct;217(4):537-539. [doi: 10.1192/bjp.2020.109] [Medline: 32423523]
24. Gautam M, Kaur M, Mahr G. COVID-19-associated psychiatric symptoms in health care workers: viewpoint from internal medicine and psychiatry residents. Psychosomatics 2020;61(5):579-581. [doi: 10.1016/j.psym.2020.04.009] [Medline: 32439184]
25. Kemp MT, Rivard SJ, Anderson S, et al. Trainee wellness and safety in the context of COVID-19: the experience of one institution. Acad Med 2021 May 1;96(5):655-660. [doi: 10.1097/ACM.0000000000003853] [Medline: 33208674]
26. West CP, Dyrbye LN, Erwin PJ, Shanafelt TD. Interventions to prevent and reduce physician burnout: a systematic review and meta-analysis. Lancet 2016 Nov 5;388(10057):2272-2281. [doi: 10.1016/S0140-6736(16)31279-X] [Medline: 27692469]
27. Sandelowski M. Whatever happened to qualitative description? Res Nurs Health 2000 Aug;23(4):334-340. [doi: 10.1002/1098-240x(200008)23:4<334::aid-nur9>3.0.co;2-g] [Medline: 10940958]
28. Sandelowski M. What's in a name? Qualitative description revisited. Res Nurs Health 2010 Feb;33(1):77-84. [doi: 10.1002/nur.20362] [Medline: 20014004]
29. Zuo KJ, Meng Y, Gordon L, et al. Navigating the postgraduate research fellowship: a roadmap for surgical residents. J Surg Res 2020 Dec;256:282-289. [doi: 10.1016/j.jss.2020.06.054] [Medline: 32712442]
30. Nguyen MA, Castelo M, Greene B, et al. Profiles of burnout and response to the COVID-19 pandemic among general surgery residents at a large academic training program. Surg Innov 2023 Apr;30(2):239-250. [doi: 10.1177/15533506221120145] [Medline: 35971874]
31. Coronavirus: what's happening in Canada and around the world on Tuesday. CBC. 2022 Feb 15. URL: https://www.cbc.ca/news/world/coronavirus-covid19-canada-world-feb15-2022-1.6352017 [accessed 2025-09-20]
32. SurveyMonkey. URL: https://www.surveymonkey.com/ [accessed 2025-11-10]
33. Dyrbye LN, Burke SE, Hardeman RR, et al. Association of clinical specialty with symptoms of burnout and career choice regret among US resident physicians. JAMA 2018 Sep 18;320(11):1114-1130 Retracted inJAMA. 2019 Mar 26;321(12):1220-1221. [doi: 10.1001/jama.2019.0167] [Medline: 30912842]
34. Shanafelt TD, Hasan O, Dyrbye LN, et al. Changes in burnout and satisfaction with work-life balance in physicians and the general US working population between 2011 and 2014. Mayo Clin Proc 2015 Dec;90(12):1600-1613. [doi: 10.1016/j.mayocp.2015.08.023] [Medline: 26653297]
35. Riley MR, Mohr DC, Waddimba AC. The reliability and validity of three-item screening measures for burnout: evidence from group-employed health care practitioners in upstate New York. Stress Health 2018 Feb;34(1):187-193. [doi: 10.1002/smi.2762] [Medline: 28524379]
36. Green J, Thorogood N. Qualitative Methods for Health Research: SAGE Publications; 2018.
37. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
38. Roulston K. Reflective Interviewing: A Guide to Theory and Practice: SAGE Publications; 2010. [doi: 10.4135/9781446288009]
39. Your contract. PARO. URL: https://myparo.ca/your-contract/ [accessed 2022-03-09]
40. Tracy BM, Stadeli KM, Coleman JR, Aribindi V, Ryan R, Lee KB. The transformation of surgical education and its influence on resident wellness. American College of Surgeons. 2019 Aug. URL: https://bulletin.facs.org/2019/08/the-transformation-of-surgical-education-and-its-influence-on-resident-wellness/ [accessed 2025-11-10]
41. Philibert I, Friedmann P, Williams WT, ACGME Work Group on Resident Duty Hours. Accreditation Council for Graduate Medical Education. New requirements for resident duty hours. JAMA 2002 Sep 4;288(9):1112-1114. [doi: 10.1001/jama.288.9.1112] [Medline: 12204081]
42. Ahmed N, Devitt KS, Keshet I, et al. A systematic review of the effects of resident duty hour restrictions in surgery. Ann Surg 2014 May 12;259(6):1041-1053. [doi: 10.1097/SLA.0000000000000595] [Medline: 24662409]
43. Gretton-Watson P, Leggat SG, Oakman J. Workplace bullying in surgery: exploring the drivers and mitigators. J Health Organ Manag 2025 Oct 13;39(6):1087-1111. [doi: 10.1108/JHOM-11-2024-0477] [Medline: 41082652]
44. Clements JM, King M, Nicholas R, et al. Bullying and undermining behaviours in surgery: a qualitative study of surgical trainee experiences in the United Kingdom (UK) & Republic of Ireland (ROI). Int J Surg 2020 Dec;84:219-225. [doi: 10.1016/j.ijsu.2020.07.031] [Medline: 32738542]

45. Munro CE, Phillips AW. Bullying in the workplace. Surgery (Oxford) 2023 Aug;41(8):516-522. [doi: 10.1016/j.mpsur.2023.05.007]
46. Lucas BP, Trick WE, Evans AT, et al. Effects of 2- vs 4-week attending physician inpatient rotations on unplanned patient revisits, evaluations by trainees, and attending physician burnout: a randomized trial. JAMA 2012 Dec 5;308(21):2199-2207. [doi: 10.1001/jama.2012.36522] [Medline: 23212497]
47. Linzer M, Poplau S, Grossman E, et al. A cluster randomized trial of interventions to improve work conditions and clinician burnout in primary care: results from the healthy work place (HWP) study. J Gen Intern Med 2015 Aug;30(8):1105-1111. [doi: 10.1007/s11606-015-3235-4] [Medline: 25724571]
48. Parshuram CS, Amaral AC, Ferguson ND, et al. Patient safety, resident well-being and continuity of care with different resident duty schedules in the intensive care unit: a randomized trial. CMAJ 2015 Mar 17;187(5):321-329. [doi: 10.1503/cmaj.140752] [Medline: 25667258]
49. Chen JH, Keyes SA, DeGregorio V, Gardner AK. Enhancing perceptions of social support and belonging in surgery residency programs. J Surg Educ 2025 Sep;82(9):103612. [doi: 10.1016/j.jsurg.2025.103612] [Medline: 40682992]
50. Rouder J, Saucier O, Kinder R, Jans M. What to do with all those open-ended responses? Data visualization techniques for survey researchers. Surv Pract 2021;14(1):1-9. [doi: 10.29115/SP-2021-0008]
51. LaDonna KA, Taylor T, Lingard L. Why open-ended survey questions are unlikely to support rigorous qualitative insights. Acad Med 2018 Mar;93(3):347-349. [doi: 10.1097/ACM.0000000000002088] [Medline: 29215376]
52. Swart R. Thematic Analysis of Survey Responses From Undergraduate Students: SAGE Publications; 2019.
53. Rich JL, Chojenta C, Loxton D. Quality, rigour and usefulness of free-text comments collected by a large population based longitudinal study - ALSWH. PLoS ONE 2013 Jul 11;8(7):e68832. [doi: 10.1371/journal.pone.0068832] [Medline: 23874784]
54. Kitto SC, Chesters J, Grbich C. Quality in qualitative research. Med J Aust 2008 Feb 18;188(4):243-246. [doi: 10.5694/j.1326-5377.2008.tb01595.x] [Medline: 18279135]
55. Afonja S, Salmon DG, Quailey SI, Lambert WM. Postdocs' advice on pursuing a research career in academia: a qualitative analysis of free-text survey responses. PLoS ONE 2021 May 6;16(5):e0250662. [doi: 10.1371/journal.pone.0250662] [Medline: 33956818]
56. Thomas DR. A general inductive approach for analyzing qualitative evaluation data. Am J Eval 2006;27(2):237-246. [doi: 10.1177/1098214005283748]
57. Why is Canada's Covid death rate so much lower than US? BBC. 2022 Feb 15. URL: https://www.bbc.com/news/world-us-canada-60380317 [accessed 2022-03-25]
58. McKim C. Meaningful member-checking: a structured approach to member-checking. Am J Qual Res 2023;7(2):41-52 [FREE Full text]

Abbreviations

ICU: intensive care unit
MBI-HSS: Maslach Burnout Inventory–Human Services Survey for Medical Personnel

Edited by R Saffary; submitted 18.02.25; peer-reviewed by J Evans, S Mallick; revised version received 23.09.25; accepted 15.10.25; published 24.11.25.

Please cite as:
Bilgen I, Castelo M, Reel E, Nguyen MA, Greene B, Lu J, Brar S, Cil T
Barriers to Wellness Among General Surgery Residents During the COVID-19 Pandemic: Qualitative Analysis of Survey Responses
JMIR Perioper Med 2025;8:e72819
URL: https://periop.jmir.org/2025/1/e72819
doi:10.2196/72819

Original Paper

Agreement Between Provider-Completed and Patient-Completed Preoperative Frailty Screening Using the Clinical Risk Analysis Index: Cross-Sectional Questionnaire Study

Mehraneh Khalighi[1,2*], MD; Amy C Thomas[3,4*], MD; Karl J Brown[4*], MS; Katherine C Ritchey[3,4*], MPH, DO

[1]Division of General Internal Medicine, Department of Medicine, University of Washington, Seattle, WA, United States

[2]Hospital and Specialty Medicine, VA Puget Sound Health Care System, Seattle, WA, United States

[3]Division of Geriatrics and Gerontology, Department of Medicine, University of Washington, Seattle, WA, United States

[4]Geriatric Research, Education and Clinical Center, VA Puget Sound Health Care System, Seattle, WA, United States

[*]all authors contributed equally

Corresponding Author:
Mehraneh Khalighi, MD
Hospital and Specialty Medicine
VA Puget Sound Health Care System
1660 South Columbian Way
S-112-SDU
Seattle, WA, 98108
United States
Phone: 1 206 764 2183
Fax: 1 206 277 6478
Email: mehraneh.khalighi@va.gov

Abstract

Background: Frailty is associated with postoperative morbidity and mortality. Preoperative screening and management of persons with frailty improves postoperative outcomes. The Clinical Risk Analysis Index (RAI-C) is a validated provider-based screening tool for assessing frailty in presurgical populations. Patient self-screening for frailty may provide an alternative to provider-based screening if resources are limited; however, the agreement between these 2 methods has not been previously explored.

Objective: The objective of our study was to examine provider-completed versus patient-completed RAI-C assessments to identify areas of disagreement between the 2 methods and inform best practices for RAI-C screening implementation.

Methods: Orthopedic physicians and physician assistants completed the RAI-C assessment on veterans aged 65 years and older undergoing elective total joint arthroplasty (eg, total hip or knee arthroplasty) and documented scores into the electronic health record during their preoperative clinic evaluation. Participants were then mailed the same RAI-C form after preoperative evaluation and returned responses to study coordinators. Agreement between provider-completed and patient-completed RAI-C assessments and differences within individual domains were compared.

Results: A total of 49 participants aged 65 years and older presenting for total joint arthroplasty underwent RAI-C assessment between November 2022 and August 2023. In total, 41% (20/49) of participants completed and returned an independent postvisit RAI-C assessment before surgery and within 180 days of their initial evaluation. There was a moderate but statistically significant correlation between provider-completed and patient-completed RAI-C assessments (r=0.62; 95% CI 0.25-0.83; P=.003). Provider-completed and patient-completed RAI-C assessments resulted in the same frailty classification in 60% (12/20) of participants, but 40% (8/20) of participants were reclassified to a more frail category based on patient-completed assessment. Agreement was the lowest between provider-completed and patient-completed screening questions regarding memory and activities of daily living.

Conclusions: RAI-C had moderate agreement when completed by providers versus the participants themselves, with more than a third of patient-completed screens resulting in a higher frailty classification. Future studies will need to explore the differences between and accuracy of RAI-C screening approaches to inform best practices for preoperative RAI-C assessment implementation.

(JMIR Perioper Med 2025;8:e66440) doi:10.2196/66440

KEYWORDS

Risk Analysis Index; preoperative screening; questionnaire; frailty; self-reported; veteran; hip; knee; arthroplasty; elective surgery; cross-sectional; quality improvement

Introduction

Frailty is a multidimensional syndrome characterized by decreased physiological reserve reducing recovery from stressors including surgery and is associated with increased postoperative morbidity and mortality [1]. Frailty screening and multidisciplinary management of persons with frailty before elective surgery improve perioperative functional performance, decrease postoperative mortality, and may improve postoperative morbidity [2,3]. While numerous patient-completed frailty screening tools (eg, FRAIL Scale, Edmonton Frail Scale, and Vulnerable Elders Survey) have been used to predict surgical morbidity and mortality in different surgical populations, few have undergone as extensive validation in the presurgical population as the Clinical Risk Analysis Index (RAI-C) [4,5]. The RAI-C is a validated 14-item health and functioning questionnaire developed to distinguish between frail and robust persons in the preoperative setting. It calculates a score between 0 and 81 from information provided by a person or surrogate with scores ≥37 indicating frailty [5-7]. Higher RAI-C scores have been associated with postoperative mortality across surgical specialties suggesting its use as an easily administered preoperative risk-stratification tool [6-8]. RAI-C has been adopted by the Veterans Health Administration (VHA) as the preferred tool for presurgical frailty assessment with the goal to optimize the care of at-risk persons [9].

Validation studies suggest that persons can complete the RAI-C independently, which is advantageous if provider time is limited [6-8]. However, review of study methods indicates that providers modified participant responses as needed, suggesting that screening was not entirely patient-led [6,7]. It is uncertain how often providers changed participant responses, which domains were modified, and how modifications affected frailty classifications. Therefore, we sought to examine provider-completed versus patient-completed RAI-C assessments to identify areas of disagreement between the 2 methods and inform best practices for RAI-C screening implementation.

Methods

Screening Procedures

Overview

As part of a quality improvement initiative, we designed and implemented a cross-sectional pilot examination to screen participants aged 65 years and older referred to an outpatient, VHA orthopedic clinic for elective total joint arthroplasty (TJA; eg, total hip or knee arthroplasty) for frailty between November 2022 and August 2023. The primary aim was to examine the agreement between provider-completed and patient-completed RAI-C assessments to inform frailty screening practices at our institution.

Orthopedic physicians and physician assistants underwent training on the use of an electronic health record (EHR)–embedded web-based RAI-C questionnaire. During preoperative evaluations, providers screened participants for frailty using the EHR-embedded assessment and recorded the RAI-C scores. Robust (RAI-C <30), prefrail (RAI-C 30-36), and frail (RAI-C ≥37) classifications were based on cutoffs defined in a large recalibration and external validation study of patients undergoing major elective noncardiac surgery. In that study, the 180-day postoperative mortality rate for RAI-C ≥30 was 2.0%, surpassing the overall mean mortality rate of 1.8%, and 4.3% for RAI-C ≥37, which is greater than twice the mean mortality rate of the population [5]. All participants also underwent screening for dementia with the Mini-Cog, a validated cognitive screen combining 3-item word memory and clock drawing [10]. Scores range from 0 to 5, with scores <3 indicating significant risk for dementia [11]. After the visit, participants were mailed a paper version of the RAI-C with a letter explaining the purpose of the screening tool and instructions on how to send back completed forms to study coordinators.

Participants were excluded from the study if surgery was performed before the participant responses to the RAI-C were received for analysis to mitigate possible confounding effects of surgery on patient-completed RAI-C responses. In addition, we excluded participant responses that were received more than 180 days from the date of provider-completed RAI-C to avoid confounding effects of progressive loss of function and osteoarthritis-related pain on the patient-completed RAI-C results. We chose an exclusion cutoff of 180 days based on findings that in individuals awaiting TJA for more than 180 days, worsening patient-reported outcome measures (ie, joint-specific function and health-related quality of life) were associated with increased levels of clinical frailty [12].

Analysis of Intervention and Measures

Patient-completed RAI-C responses were compared with provider-completed EHR-RAI-C results and analyzed for discrepancies between their total RAI-C and individual domain scores. The study authors performed a detailed EHR review to verify accuracy of provider and participant responses pertaining to health conditions (ie, presence of renal failure, heart failure, weight loss, or cancer). The provider completing the RAI-C also performed Mini-Cog screening for dementia to identify persons who would benefit from geriatric consultation (eg, scores <3), but results of this screening did not inform the subjective participant responses to the RAI-C question on loss of memory. The accuracy of participant responses to subjective questions (ie, limitations in activities of daily living [ADLs], loss of appetite, or memory problems) was not verified.

The primary outcome measure was the degree of concordance between provider-completed and patient-completed total RAI-C scores. Secondary outcome measures were degree of concordance between the responses for individual domains and

the effect of time elapsed between provider-completed and patient-completed responses on the degree of concordance between scores. The Pearson product-moment correlation coefficient (r) was used to determine the linear relationship between provider-completed and patient-completed total RAI-C and individual RAI-C domain scores and time elapsed in days from provider to participant completion of the RAI-C and the absolute difference in scores obtained, respectively. Quantile-quantile plots and histograms of both the provider-completed and patient-completed total RAI-C scores indicated that the distributions of both variables were approximately normal. All analyses were performed in R (version 4.3.1; R Foundation for Statistical Computing).

Ethical Considerations

The Human Research Protection Program, Associate Chief of Staff for Research and Development, and Quality, Safety, and Values department reviewed this project in accordance with the Veterans Health Administration Program Guide 1200.21 and determined that it was a nonresearch, operations activity. Thus, approval by an institutional review board and consent to participate were not needed. Participant data were anonymized to ensure privacy and confidentiality. Participants were not offered compensation.

Results

Forty-nine participants aged 65 years and older presenting for TJA underwent RAI-C screening between November 2022 and August 2023. In total, 61% (30/49) of participants returned a postvisit RAI-C assessment, but 9 participants underwent surgery before completion and were excluded from analysis. An additional participant who returned a postvisit RAI-C assessment more than 180 days from orthopedic clinic evaluation was excluded. Therefore, 41% (20/49) of participants who returned a completed postvisit RAI-C assessment before surgery within 180 days from their initial evaluation were included in our analysis and their characteristics are summarized in Table 1. The number of positive responses to RAI-C questions reported in Table 2 show all responses. Identical result counts between provider and patient responses do not necessarily indicate agreement between their respective responses.

We used RAI-C score without cancer in our analysis since none of the participants met RAI-C definition of cancer (ie, unresectable cancer, metastatic cancer with poor prognosis, chemotherapy within 30 days, or radiotherapy within 90 days). There was statistically significant, moderate correlation between provider-completed and patient-completed RAI-C (N=20, r=0.62, 95% CI 0.25-0.83; P=.003; Figure 1).

Table 1. Participant characteristics (N=20).

Characteristics	Values
Gender, n (%)	
Men	19 (95)
Average age, years (range)	74 (66-83)
Race, n (%)[a]	
White	17 (85)
Black	2 (10)
Preferred language, n (%)[b]	
English	18 (90)
Mini-Cog score ≥3, n (%)	19 (95)

[a]One participant declined to respond.

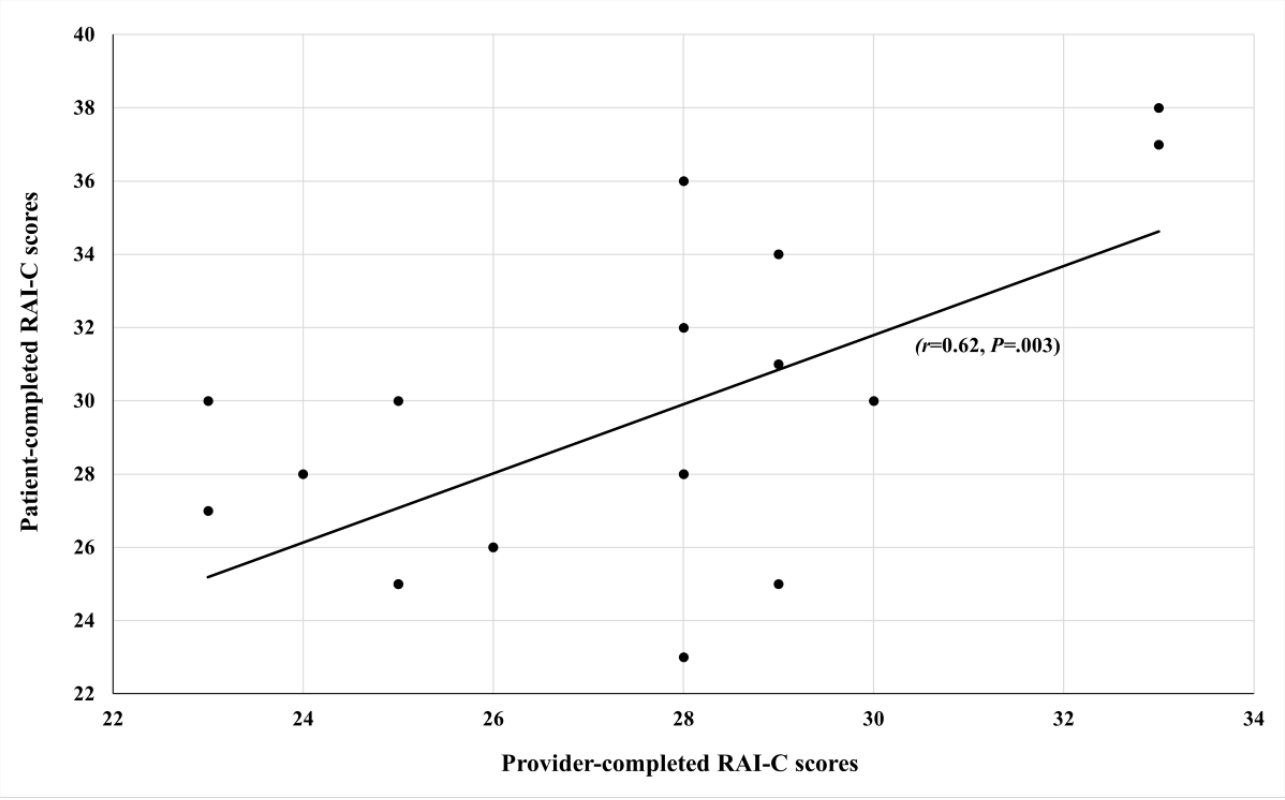
[b]Two participants declined to respond.

Table 2. Patients' and providers' responses.

Factors	Patient-completed	Provider-completed
Medical conditions per RAI-C[a] definition, n (%)		
Kidney disease	0 (0)	0 (0)
Heart failure	3 (15)	0 (0)
Shortness of breath	0 (0)	0 (0)
Cancer within 5 years	0 (0)	0 (0)
Nutrition, n (%)		
Loss of weight	3 (15)	1 (5)
Loss of appetite	1 (5)	0 (0)
Cognition, n (%)		
Loss of memory	4 (20)	3 (15)
Limitations in activities of daily living, n (%)		
Mobility	10 (50)	10 (50)
Eating	3 (15)	1 (5)
Toileting	2 (10)	0 (0)
Personal hygiene	2 (10)	0 (0)
Total RAI-C score, n (%)		
RAI-C <30 (Robust)	11 (55)	17 (85)
RAI-C 30-36 (Prefrail)	7 (35)	3 (15)
RAI-C ≥ 37 (Frail)	2 (10)	0 (0)

[a]RAI-C: Clinical Risk Analysis Index.

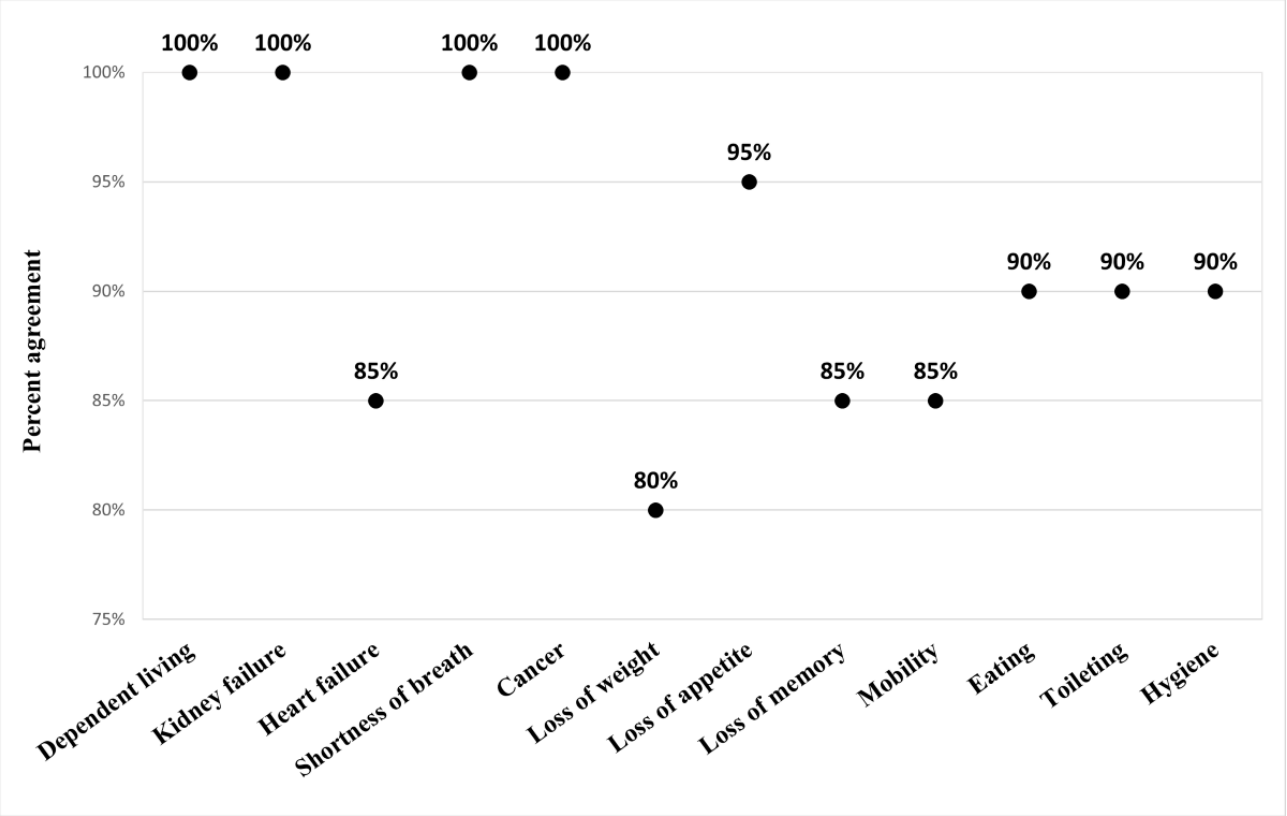
Figure 1. Correlation between provider-completed and patient-completed total RAI-C scores (N=20). RAI-C: Clinical Risk Analysis Index.



Frailty classification was identical in 60% (12/20) of participants. The remaining 40% (8/20) of participants were reclassified to a higher level of frailty based on patient-completed RAI-C scores. In addition, 30% (6/20) of participants were reclassified from robust to prefrail and 10% (2/20) from prefrail to frail (Multimedia Appendix 1).

Agreement between questions concerning chronic health conditions such as kidney disease and cancer was relatively high (Figure 2).

Figure 2. Percentage agreement between provider-completed and patient-completed responses to individual domains of the Clinical Risk Analysis Index (N=20).



The domains with lowest agreements included heart failure, loss of weight, loss of memory, and the mobility subcategory of ADLs. Neither participant nor provider responses to weight loss (ie, loss of ≥10 lb in the past 3 months without trying) were accurate as they were not supported by EHR-documented weights. Although participant responses to weight loss compared with provider responses differed in 20% (4/20) of participants, this disagreement did not affect their respective frailty classification.

In responding to questions on loss of appetite, loss of memory, and limitations in ADLs, 45% (9/20) of participants assigned lower scores than providers, which reclassified 6 of these participants to a higher level of frailty. Therefore, participant responses to questions pertaining to loss of appetite, loss of memory, and ADLs accounted for 75% (6/8) of observed reclassifications to a higher level of frailty. The remaining 2 observed reclassifications to a higher level of frailty were based on participant responses indicating presence of heart failure, which was supported on review of EHR documentation of heart failure symptoms or consistent findings on transthoracic echocardiography.

On average, participants returned self-assessments within 41 days of the date the forms were mailed to them (median 28, range 21-68 days) with an average time between completion of provider and participant RAI-C forms of 65 (median 65, range 25-118) days. Time elapsed between assessments did not correlate with the differences observed between RAI-C scores (N=20, r=0.38; P=.10).

Discussion

Principal Findings

The RAI-C is preferred for preoperative frailty screening in VHA and has been validated in presurgical populations [5-7]. These prior validation studies have not fully explored the relationship between provider and patient-completed assessments as a method to increase screening efficiency. We showed that our population of older veterans with low concern for cognitive impairment presenting for elective orthopedic TJA could complete RAI-C assessments independently. However, the correlation between provider-completed and patient-completed RAI-C scores was only moderate and more than a third of participants were reclassified to higher levels of frailty based on self-assessment. While other studies comparing provider versus participant perceptions of frailty also observed moderate correlation between the 2 methods, their study populations and settings were different (emergency room vs preoperative setting), they used a different screening tool (Clinical Frail Scale vs RAI-C), and they found that providers

assigned higher levels of frailty than participants [13,14]. Our study is one of the first to highlight areas of discrepancy between provider-completed and patient-completed RAI-C, suggesting challenges to the predictive validity of this tool and considerations for clinical implementation.

We found that disagreement between provider and participant responses and reclassifications were mainly based on participant-perceived decline in appetite, memory, and performance of ADLs, or heart failure. Notably, all participants accurately recognized their heart failure diagnosis, while providers missed the diagnosis in 3 cases. Disagreement between provider and participant responses to these domains (ie, heart failure, loss of appetite, memory loss, and limitations in ADLs) and provider underclassification of frailty has potentially significant clinical ramifications. Although optimal management of frailty is ill-defined, expert consensus suggests that persons with frailty should undergo comprehensive assessments to identify and address rehabilitative, nutritional, and psychosocial needs preoperatively [15]. Emerging data suggest that multimodal interventions can improve postsurgical outcomes for persons with frailty undergoing elective surgeries [3,16,17]. High-risk surgical candidates with frailty should have exploration of their health care priorities, postsurgical goals, and care preferences to avoid potentially deleterious postoperative outcomes [18]. Clarification of goals of care in the context of surgical risk and expected clinical outcomes, termed “surgical pause,” increases receipt of goal-concordant care and avoids unwanted surgery [19]. Thus, adequately and accurately identifying level of functional ability, cognition, and ultimately frailty of preoperative persons is important for unbiased care planning and resource allocation.

However, disuse or incorrect use of frailty screening tools can contribute to misclassification of frailty, potentially limiting access to interventions and significantly impacting quality of life and function. Elective TJA is rarely lifesaving but significantly impacts functional ability and preservation of independence [20]. Without consistent use of validated tools to screen for frailty, ageism and other implicit biases may contribute to overclassification of frailty by health care professionals and increase their reluctance to offer therapies simply based on biological age or “old” appearance [21]. Alternatively, concerns about surgical candidacy, unaddressed pain, and further loss of function may contribute to social desirability and response biases that encourage underclassification of frailty by participants who are reluctant to report functional or other limitations when responding to provider questions assessing for presurgical frailty [22,23]. Similar to responses to sensitive questions, where perceptions of anonymity and privacy increase the accuracy of self-reported answers, written responses to questions on performance of ADLs may be more accurate than verbal responses to providers, especially during the first encounter when participants have not yet built rapport with their providers [24,25].

In addition, the lived experiences of older adults and their perception of health may influence frailty classification and related health outcomes [26]. The person’s perception of decline in one domain (eg, performance of ADLs) may affect performance in other domains (eg, decline in appetite or memory) with a cumulative effect on level of frailty [26]. Therefore, the participants’ responses could be considered a more accurate reflection of subjective symptoms or functional ability, as they represent the individuals’ perceptions of their health.

When participants respond to the same questions without provider oversight, the effect of these biases may be minimized, and the accuracy of the screening tool might improve.

Limitations

Our evaluation was limited to a small population of mainly English-speaking men with low concern for cognitive impairment within 1 VHA orthopedic surgery clinic which may not relate to other presurgical populations (eg, peripheral vascular surgery or general surgery) with different prevalences of frailty and cognitive impairment. In addition, worsening joint-specific function and health-related quality of life with longer wait times before TJA or surgical intervention between provider and patient-completed RAI-C can influence participants’ responses. Therefore, we attempted to mitigate possible confounding effects of prolonged wait times before surgery by excluding participant-completed RAI-C results that were completed more than 180 days from provider-completed surveys. We attempted to mitigate the effect of surgery on patient-completed RAI-C by excluding those participants who underwent surgery before completing the self-reported RAI-C. Nonetheless, our study was strengthened by the high participant response rate of more than 40%. In most cases of disagreement (ie, cognition and limitations in ADLs), participants’ responses resulted in a higher frailty classification, which could not be verified for accuracy. Furthermore, we could not assess for the role of selection bias on our findings. It is possible that self-reported responses to the RAI-C were predominantly completed and returned by participants who disagreed with provider-completed responses to the RAI-C. Finally, participant completion of the RAI-C relies on the ability to read and understand the questions. We were unable to assess the effects of health literacy or educational level on assessment disagreements.

Conclusions

Frailty screening with the RAI-C can be done by providers or patients before elective orthopedic TJA. The level of disagreement observed between provider-completed and patient-completed assessments suggests that these methods are not interchangeable. Future studies exploring screening methods in larger, more diverse populations who are undergoing a variety of surgeries may clarify challenges to screening accuracy and validity of patient-completed screening approaches.

Acknowledgments

Partial financial support was received from Veterans Integrated Service Network (VISN) 20 for clinical demonstration and innovation programs.

Authors' Contributions

MK, AT, and KR contributed to the concept and design of the project; MK and KR took the lead in drafting the manuscript; and MK and KB conducted the data analysis. All authors provided critical feedback and helped shape this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Frailty classification based on provider-completed versus patient-completed RAI-C scores (N=20). RAI-C: Clinical Risk Analysis Index.
[PNG File , 51 KB - periop_v8i1e66440_app1.png]

References

1. Shinall MC, Arya S, Youk A, Varley P, Shah R, Massarweh NN, et al. Association of preoperative patient frailty and operative stress with postoperative mortality. JAMA Surg 2020;155(1):e194620 [FREE Full text] [doi: 10.1001/jamasurg.2019.4620] [Medline: 31721994]
2. Hall DE, Arya S, Schmid KK, Carlson MA, Lavedan P, Bailey TL, et al. Association of a frailty screening initiative with postoperative survival at 30, 180, and 365 Days. JAMA Surg 2017;152(3):233-240 [FREE Full text] [doi: 10.1001/jamasurg.2016.4219] [Medline: 27902826]
3. Hall DE, Youk A, Allsup K, Kennedy K, Byard TD, Dhupar R, et al. Preoperative rehabilitation is feasible in the weeks prior to surgery and significantly improves functional performance. J Frailty Aging 2023;12(4):267-276 [FREE Full text] [doi: 10.14283/jfa.2022.42] [Medline: 38008976]
4. Nidadavolu LS, Ehrlich AL, Sieber FE, Oh ES. Preoperative evaluation of the frail patient. Anesth Analg 2020;130(6):1493-1503 [FREE Full text] [doi: 10.1213/ANE.0000000000004735] [Medline: 32384339]
5. Arya S, Varley P, Youk A, Borrebach JD, Perez S, Massarweh NN, et al. Recalibration and external validation of the risk analysis index: a surgical frailty assessment tool. Ann Surg 2020;272(6):996-1005 [FREE Full text] [doi: 10.1097/SLA.0000000000003276] [Medline: 30907757]
6. Hall DE, Arya S, Schmid KK, Blaser C, Carlson MA, Bailey TL, et al. Development and initial validation of the risk analysis index for measuring frailty in surgical populations. JAMA Surg 2017;152(2):175-182 [FREE Full text] [doi: 10.1001/jamasurg.2016.4202] [Medline: 27893030]
7. Shah R, Borrebach JD, Hodges JC, Varley PR, Wisniewski MK, Shinall MC, et al. Validation of the risk analysis index for evaluating frailty in ambulatory patients. J Am Geriatr Soc 2020;68(8):1818-1824 [FREE Full text] [doi: 10.1111/jgs.16453] [Medline: 32310317]
8. George EL, Hall DE, Youk A, Chen R, Kashikar A, Trickey AW, et al. Association between patient frailty and postoperative mortality across multiple noncardiac surgical specialties. JAMA Surg 2021;156(1):e205152 [FREE Full text] [doi: 10.1001/jamasurg.2020.5152] [Medline: 33206156]
9. The surgical pause practice adopted as national program. US Department of Veterans Affairs. 2024 Feb 13. URL: https://www.hsrd.research.va.gov/impacts/surgical-pause.cfm [accessed 2024-07-13]
10. Borson S, Scanlan JM, Chen P, Ganguli M. The mini-cog as a screen for dementia: validation in a population-based sample. J Am Geriatr Soc 2003;51(10):1451-1454 [FREE Full text] [doi: 10.1046/j.1532-5415.2003.51465.x] [Medline: 14511167]
11. Borson S, Scanlan J, Brush M, Vitaliano P, Dokmak A. The mini-cog: a cognitive 'vital signs' measure for dementia screening in multi-lingual elderly. Int J Geriatr Psychiatry 2000;15(11):1021-1027. [doi: 10.1002/1099-1166(200011)15:11<1021::aid-gps234>3.0.co;2-6] [Medline: 11113982]
12. Bayram JM, Wickramasinghe NR, Scott CEH, Clement ND. Clinical frailty is independently associated with joint-specific function and health-related quality of life in patients awaiting a total hip or knee arthroplasty. Bone Jt Open 2023 Apr 7;4(4):241-249. [doi: 10.1302/2633-1462.44.bjo-2023-0020.r1]
13. Shah SP, Penn K, Kaplan SJ, Vrablik M, Jablonowski K, Pham TN, et al. Comparison of bedside screening methods for frailty assessment in older adult trauma patients in the emergency department. Am J Emerg Med 2019;37(1):12-18. [doi: 10.1016/j.ajem.2018.04.028] [Medline: 29728285]
14. Dresden SM, Platts-Mills TF, Kandasamy D, Walden L, Betz ME. Patient versus physician perceptions of frailty: a comparison of clinical frailty scale scores of older adults in the emergency department. Acad Emerg Med 2019;26(9):1089-1092 [FREE Full text] [doi: 10.1111/acem.13825] [Medline: 31265194]

15. Alvarez-Nebreda ML, Bentov N, Urman RD, Setia S, Huang JC, Pfeifer K, et al. Recommendations for preoperative management of frailty from the Society for Perioperative Assessment and Quality Improvement (SPAQI). J Clin Anesth 2018;47:33-42. [doi: 10.1016/j.jclinane.2018.02.011] [Medline: 29550619]
16. Pouwels S, Stokmans RA, Willigendael EM, Nienhuijs SW, Rosman C, van Ramshorst B, et al. Preoperative exercise therapy for elective major abdominal surgery: a systematic review. Int J Surg 2014;12(2):134-140 [FREE Full text] [doi: 10.1016/j.ijsu.2013.11.018] [Medline: 24325942]
17. Cabilan CJ, Hines S, Munday J. The effectiveness of prehabilitation or preoperative exercise for surgical patients: a systematic review. JBI Database System Rev Implement Rep 2015;13(1):146-187. [doi: 10.11124/jbisrir-2015-1885] [Medline: 26447015]
18. Reuben DB, Jennings LA. Putting goal-oriented patient care into practice. J Am Geriatr Soc 2019;67(7):1342-1344 [FREE Full text] [doi: 10.1111/jgs.15885] [Medline: 30882888]
19. Oyekan AA, Lee JY, Hodges JC, Chen SR, Wilson AE, Fourman MS, et al. Increasing quality and frequency of goals-of-care documentation in the highest-risk surgical candidates: one-year results of the surgical pause program. JB JS Open Access 2023;8(2):e22.00107 [FREE Full text] [doi: 10.2106/JBJS.OA.22.00107] [Medline: 37101601]
20. Neuprez A, Neuprez AH, Kaux JF, Kurth W, Daniel C, Thirion T, et al. Total joint replacement improves pain, functional quality of life, and health utilities in patients with late-stage knee and hip osteoarthritis for up to 5 years. Clin Rheumatol 2020;39(3):861-871 [FREE Full text] [doi: 10.1007/s10067-019-04811-y] [Medline: 31720892]
21. Chang ES, Kannoth S, Levy S, Wang SY, Lee JE, Levy BR. Global reach of ageism on older persons' health: a systematic review. PLoS One 2020;15(1):e0220857 [FREE Full text] [doi: 10.1371/journal.pone.0220857] [Medline: 31940338]
22. Bispo JP. Social desirability bias in qualitative health research. Rev Saude Publica 2022;56:101 [FREE Full text] [doi: 10.11606/s1518-8787.2022056004164] [Medline: 36515303]
23. Tay Swee Cheng R, Klainin-Yobas P, Hegney D, Mackey S. Factors relating to perioperative experience of older persons undergoing joint replacement surgery: an integrative literature review. Disabil Rehabil 2015;37(1):9-24. [doi: 10.3109/09638288.2014.906663] [Medline: 24689440]
24. Durant LE, Carey MP. Self-administered questionnaires versus face-to-face interviews in assessing sexual behavior in young women. Arch Sex Behav 2000;29(4):309-322. [doi: 10.1023/a:1001930202526] [Medline: 10948721]
25. Ong AD, Weiss DJ. The Impact of Anonymity on Responses to Sensitive Questions. J Applied Social Pyschol 2006 Jul 31;30(8):1691-1708. [doi: 10.1111/j.1559-1816.2000.tb02462.x]
26. Khalil AH, Gobbens RJJ. What If the clinical and older adults' perspectives about frailty converge? A call for a mixed conceptual model of frailty: a traditional literature review. Healthcare (Basel) 2023;11(24):3174 [FREE Full text] [doi: 10.3390/healthcare11243174] [Medline: 38132064]

Abbreviations

ADLs: activities of daily living
EHR: electronic health record
RAI-C: Clinical Risk Analysis Index
TJA: total joint arthroplasty
VHA: Veterans Health Administration

Edited by J Shiffermiller; submitted 12.09.24; peer-reviewed by J Hejkal, V Aznar-Tortonda; comments to author 24.10.24; revised version received 13.12.24; accepted 17.01.25; published 10.02.25.

Please cite as:
Khalighi M, Thomas AC, Brown KJ, Ritchey KC
Agreement Between Provider-Completed and Patient-Completed Preoperative Frailty Screening Using the Clinical Risk Analysis Index: Cross-Sectional Questionnaire Study
JMIR Perioper Med 2025;8:e66440
URL: https://periop.jmir.org/2025/1/e66440
doi:10.2196/66440
PMID:39928399

Evaluating Large Language Models for Preoperative Patient Education in Superior Capsular Reconstruction: Comparative Study of Claude, GPT, and Gemini

Yukang Liu[1*], MBBS; Hua Li[2*], PhD; Jianfeng Ouyang[3*], PhD; Zhaowen Xue[4], PhD; Min Wang[5], PhD; Hebei He[4], PhD; Bin Song[6], PhD; Xiaofei Zheng[4], PhD; Wenyi Gan[3*], PhD

[1]The Second School of Clinical Medicine, Southern Medical University, Guangzhou, China

[2]Department of Orthopedics, Beijing Jishuitan Hospital, Beijing, China

[3]Zhuhai People's Hospital (The Affiliated Hospital of Beijing Institute of Technology, Zhuhai Clinical Medical College of Jinan University), 79 Kangning Road, Xiangzhou District, Zhuhai, Guangdong, China

[4]Department of Sports Medicine, The First Affiliated Hospital of Jinan University, Guangzhou, China

[5]Department of Orthopaedics, Guangzhou Red Cross Hospital of Jinan University, Guangzhou, China

[6]Department of Joint Surgery and Sports Medicine, The Sixth Affiliated Hospital of Sun Yat-sen University, Guangzhou, China

[*]these authors contributed equally

Corresponding Author:
Wenyi Gan, PhD
Zhuhai People's Hospital (The Affiliated Hospital of Beijing Institute of Technology, Zhuhai Clinical Medical College of Jinan University), 79 Kangning Road, Xiangzhou District, Zhuhai, Guangdong, China

Abstract

Background: Large language models (LLMs) are revolutionizing natural language processing, increasingly applied in clinical settings to enhance preoperative patient education.

Objective: This study aimed to evaluate the effectiveness and applicability of various LLMs in preoperative patient education by analyzing their responses to superior capsular reconstruction (SCR)–related inquiries.

Methods: In total, 10 sports medicine clinical experts formulated 11 SCR issues and developed preoperative patient education strategies during a webinar, inputting 12 text commands into Claude-3-Opus (Anthropic), GPT-4-Turbo (OpenAI), and Gemini-1.5-Pro (Google DeepMind). A total of 3 experts assessed the language models' responses for correctness, completeness, logic, potential harm, and overall satisfaction, while preoperative education documents were evaluated using DISCERN questionnaire and Patient Education Materials Assessment Tool instruments, and reviewed by 5 postoperative patients for readability and educational value; readability of all responses was also analyzed using the cntext package and py-readability-metrics.

Results: Between July 1 and August 17, 2024, sports medicine experts and patients evaluated 33 responses and 3 preoperative patient education documents generated by 3 language models regarding SCR surgery. For the 11 query responses, clinicians rated Gemini significantly higher than Claude in all categories (P<.05) and higher than GPT in completeness, risk avoidance, and overall rating (P<.05). For the 3 educational documents, Gemini's Patient Education Materials Assessment Tool score significantly exceeded Claude's (P=.03), and patients rated Gemini's materials superior in all aspects, with significant differences in educational quality versus Claude (P=.02) and overall satisfaction versus both Claude (P<.01) and GPT (P=.01). GPT had significantly higher readability than Claude on 3 R-based metrics (P<.01). Interrater agreement was high among clinicians and fair among patients.

Conclusions: Claude-3-Opus, GPT-4-Turbo, and Gemini-1.5-Pro effectively generated readable presurgical education materials but lacked citations and failed to discuss alternative treatments or the risks of forgoing SCR surgery, highlighting the need for expert oversight when using these LLMs in patient education.

(JMIR Perioper Med 2025;8:e70047) doi:10.2196/70047

KEYWORDS

superior capsular reconstruction; massive rotator cuff tear; large language models; preoperative patient education; informed consent process

Introduction

Large language models (LLMs) are extensive neural network models based on deep learning [1,2]. These models learn the grammar, semantics, and contextual information of a language by training on vast amounts of textual data, enabling them to perform various natural language processing tasks [1,2]. Due to the powerful text processing, text generation capabilities, and immense knowledge training of LLMs, researchers have begun to continually explore the potential of LLMs in clinical application scenarios, including professional licensing examinations in various countries and regions [3-5], answering public health questions [6,7], analyzing radiological images [8], disease screening [9], disease diagnosis [10], and discipline education [11]. As the versions and functions of LLMs are constantly updated and upgraded, these models have a low usage threshold and are convenient to use. It is particularly important for professionals in various disciplines to assess the accuracy and completeness of LLMs in their respective fields. This assessment not only provides a strong basis for the application of LLMs in various disciplines but also identifies their shortcomings, serving as a warning for nonprofessional users [3,8,10,11].

Superior capsular reconstruction (SCR) was initially proposed by Mihata et al [12] in 2012 as a technique to restore the superior restraint of the humeral head passively, thereby restoring force couples and improving shoulder joint kinematics. Over the past decade, SCR has become one of the commonly used treatment methods for massive and irreparable rotator cuff tears among clinicians [13,14]. However, the surgical techniques for SCR are highly variable [15]. For example, contrary to the results of earlier studies, further research suggests using dermal allograft instead of fascia lata autograft, leading to a current lack of sufficiently effective long-term follow-up data with high levels of evidence [16-18]. Moreover, as SCR is a reconstructive surgery rather than a repair surgery [15], it is challenging to provide patients with a standardized and effective explanation and communication during the preoperative informed consent process. An effective preoperative informed consent process is one of the essential steps in alleviating patients' perioperative anxiety and improving treatment efficacy [19,20].

Rational and effective preoperative patient education is one of the critical components in developing standardized diagnosis and treatment processes for clinical surgery departments [21]. The main difficulty lies in the professional knowledge gap between medical staff and patients [22]. Previous studies have shown that using multimedia as patient education materials can better help patients understand surgical procedures and alleviate perioperative anxiety [23,24]. However, in most cases, doctors still primarily use verbal responses to address patients' individualized questions [25]. This might probably because preparing personalized educational materials and providing oral education requires a significant investment of time and effort, leading to high time and economic costs. Furthermore, there is a vast difference in the sources of medical information accessed by doctors and patients [26]. Doctors primarily obtain medical information from clinical guidelines, research literature, and textbooks, while patients often acquire medical information through simple search engines and social media software, which may contain false and overly embellished content [26-28]. Patients often lack the ability to think independently when faced with this information.

With the development of LLMs in recent years, researchers have discovered that the disciplinary knowledge possessed by these LLMs can pass professional examinations in multiple disciplines [3,10,29]. Their powerful text processing capabilities not only allow them to polish complex text content to enhance readability but also enable them to independently generate text content that is more comprehensive and empathetic compared to health care professionals [6,7,30]. The quality of their answers is also significantly better than the search results from search engines [27,28]. Researchers have also pointed out that when using LLMs as patient education assistive tools, the primary task of doctors is to determine the accuracy of the information and make necessary clarifications [5,31]. Furthermore, researchers believe that LLMs can present information in a way that is understandable to most patients, making them a valuable supplement for orthopedic surgeons in obtaining informed consent and shared decision-making [4,5].

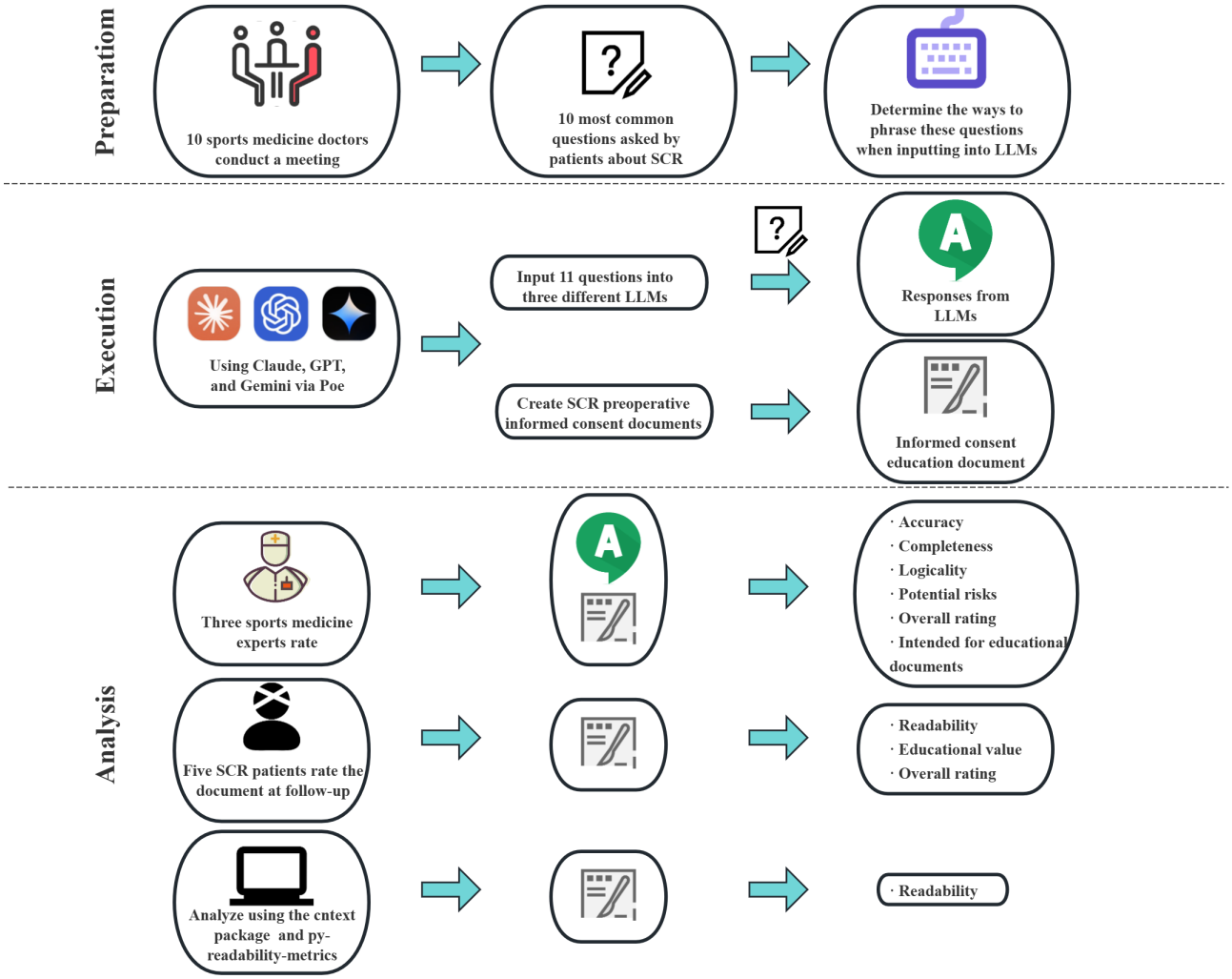
This cross-sectional study aims to assess the capability and application potential of different LLMs in preoperative patient education by evaluating the responses of 3 LLMs—GPT-4-Turbo, Claude-3-Opus, and Gemini-1.5-Pro—to SCR-related patient inquiries. In addition, the study will evaluate patient education documents generated by the LLMs for the informed consent process, which will be jointly assessed by health care professionals and patients. We hypothesize that LLMs can generate readable patient education materials for SCR, but the accuracy, completeness, and patient-assessed readability of the content will require expert review before clinical application.

Methods

Study Design Overview

This cross-sectional analysis, conducted from July 1 to August 17, 2024, evaluated the quality of responses generated by different LLMs in the context of preoperative patient education for SCR. The study design assessed Claude-3-Opus, GPT-4-Turbo, and Gemini-1.5-Pro (accessed via Poe) on their ability to answer SCR-related patient questions and generate educational materials. The specific study flow is shown in Figure 1. All LLM prompts and responses, as well as expert and patient evaluations, were conducted in Chinese. Screenshots of Poe website operations are available in Mendeley (Mendeley Data, V1), with English translations generated by GPT-4-Turbo (via Poe) in Multimedia Appendix 1.

Figure 1. Flow diagram of the study process. LLM: large language model; SCR: superior capsular reconstruction.



Ethical Considerations

This study was approved by the Ethics Committee of our organization and was eligible for exemption from ethical review considering that this cross-sectional study involved no interventions or potential risks to patients.

Questions and Prompts Development

The research team for this study consists of 12 members, including 10 experienced sports medicine clinicians and 2 doctoral students specializing in LLMs, who collaborated to create patient education materials about SCR. The clinicians include 3 senior-level experts (2 of whom are subject matter experts from external institutions), 2 associate senior-level experts, and 5 intermediate-level experts, with each clinician having at least 5 years of clinical experience.

The 2 doctoral students first collected a total of 100 questions by having each of the 10 clinical experts propose 10 questions daily that patients frequently asked about SCR, covering aspects like etiology, treatment principles, methods, complications, rehabilitation, and hospitalization costs. After removing duplicates and combining some of the questions, they included only the effective questions that all experts agreed were meaningful. This process resulted in the inclusion of 11 questions. Along with these questions, the doctoral students provided instructions (Table 1) requiring LLMs to draft a standardized preoperative informed consent patient education document. After the drafted prompts were reviewed and approved by the aforementioned 10 clinical experts, doctoral students created standardized prompts for each question, consisting of unified “Background+ Question” formats (Table 1). These standardized prompts were then used to generate a comprehensive patient education document addressing most concerns of SCR patients using LLMs.

Table . Content and strategies for asking questions to large language models.

Subject	Theme	Content
Background	Clinical case	The patient was diagnosed with a massive rotator cuff tear due to supraspinatus muscle injury. The doctor plans to perform a superior capsular reconstruction surgery on the shoulder joint.
Question 1	Muscle injury	The imaging report says that I have a supraspinatus muscle injury. What is the supraspinatus muscle, and what causes this type of injury?
Question 2	Surgical principles and indications	What is the reconstruction of the superior capsule of the shoulder joint, what is the therapeutic principle of the surgery, and what are the indications for the surgery?
Question 3	Graft materials	What are the commonly used graft materials in the reconstruction of the superior capsule of the shoulder joint, and what are the differences between these grafts?
Question 4	Surgical hardware	Besides grafts, does the reconstruction of the superior capsule of the shoulder joint require the use of screws, and do these screws need to be removed in a second surgery?
Question 5	Surgical complications	What are the surgical complications of superior capsule reconstruction of the shoulder joint?
Question 6	Recovery time	How long is the typical recovery time after superior capsule reconstruction surgery of the shoulder joint?
Question 7	Healing issues	What situations can lead to poor healing or failure of the superior capsule reconstruction surgery of the shoulder joint?
Question 8	Autograft risks	In superior capsule reconstruction surgery of the shoulder joint, if an autograft is chosen, what are the impacts and risks to the area from which the autologous tissue is harvested?
Question 9	Surgical costs	What are the chargeable items during the superior capsule reconstruction surgery of the shoulder joint, and what surgical consumables are needed?
Question 10	Graft longevity	If the superior capsule reconstruction surgery of the shoulder joint is successful, how long is the lifespan of the implanted graft, and what are the differences between different types of grafts?
Question 11	Anesthesia and hospitalization	What type of anesthesia is required for superior capsule reconstruction surgery, how long does the surgery take, and how long is the hospital stay required?
Document generation request	Education document	Please generate a comprehensive educational document about superior capsule reconstruction surgery of the shoulder joint. This document is to be provided to patients for reading during the preoperative informed consent process.

LLM Selection and Prompt Execution

Both ChatGPT 4 and Claude 3 are among the most popular language models today, with Gemini (formerly known as Bard) also gaining significant traction [32]. Studies suggest potential discrepancies in the functionalities of GPT-4 models used on the OpenAI official website [33]. To mitigate potential systematic errors arising from these discrepancies, we access Claude-3-Opus, GPT-4-Turbo, and Gemini-1.5-Pro through the Poe website. Poe, created by Anthropic, is a platform that aggregates multiple AI chatbots, enabling users to engage with different AI assistants within a single interface and compare their responses [34].

To ensure that each interaction is independent and unbiased by previous exchanges, the doctoral students perform a “clear context” operation after each query. This approach ensures that

each question and response are treated independently, preventing information carryover from previous interactions, and is informed by other research [7,11]. Since the purpose of our study was to evaluate the ability of pretrained LLMs to handle new tasks, we used LLMs in Zero-shot mode. Before input, the generated content has no specific setting (ie, suppose you are a doctor or speak like a doctor). The input provided to the LLMs follows a "background+ question/request" format (human message) and the output answers (assistant message) were collected then, ensuring clarity and relevance within each independent interaction.

Evaluation of LLM Response Quality

This study evaluates the quality of patient informed consent documents generated by LLMs from 3 perspectives: physicians' assessment, patients' assessment, and readability analysis.

In total, 3 senior doctors evaluated the LLMs' responses to 11 specific questions related to a specific medical procedure, assessing them for correctness, completeness, logic, and potential harm using a 5-point Likert scale [35]. Physicians also provided an overall satisfaction score using a 10-point Likert scale. In addition, to evaluate the quality of health care information provided by each LLM, 2 validated instruments were also used to assess the generated documents: DISCERN (score ranging from 1=low to 5=high for overall information quality) and the Patient Education Materials Assessment Tool (PEMAT) for printable materials (scores of 0% - 100% for understandability) [6]. The PEMAT assessment tool was able to assess printable and audiovisual understandability, while the DISCERN instrument could review the quality of information for the consumer particularly with a focus on treatment choices in health information.

In total, 5 patients who underwent the specific medical procedure reviewed the LLM-generated patient education documents, rating their readability and educational value on a 5-point Likert scale and overall satisfaction on a 10-point Likert scale. This aimed to assess the documents' clarity and educational value from nonprofessional readers' perspectives.

Finally, a readability analysis of all LLMs' responses was conducted using the cntext package [36] in R (version 4.4.1), examining sentence structure and evaluating readability via 3 indices: readability 1 (average characters per clause), readability 2 (proportion of adverbs and conjunctions), and readability 3, based on the Fog Index and calculated as half the sum of readability 1 and readability 2. Besides, we also applied the "py-readability-metrics" to evaluate the readability, which includes metrics such as the Flesch Reading Ease Score, Flesch-Kincaid Grade Level, and Gunning Fog Index.

Data Analysis

Statistical analysis used SPSS (version 26.0; IBM Corp) using nonparametric tests due to nonnormally distributed data (Kolmogorov-Smirnov test). Mann-Whitney *U* test compared scoring between groups, with significance at $P<.05$. Interrater reliability, assessed using Fleiss kappa value, was interpreted as follows: poor agreement (<0.01); slight agreement (0.01 - 0.20); fair agreement (0.21 - 0.40); moderate agreement (0.41 - 0.60); substantial agreement (0.61 - 0.80); almost perfect agreement (0.81 - 1.00) [7]. GraphPad Prism 8 generated bar charts for visualizing results.

Results

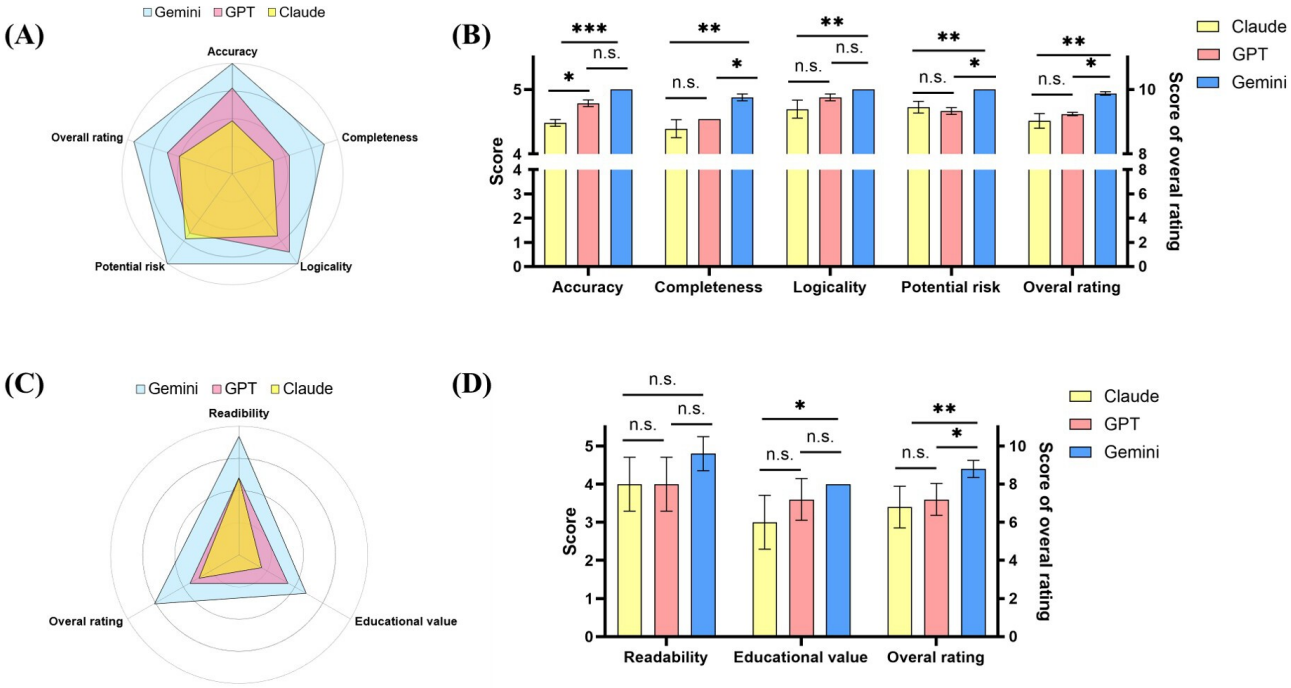
Overview

Between July 1 and July 14, 2024, we sent invitations to sports medicine experts at various hospitals in the South China region for a webinar held on July 18. During this meeting, we discussed 11 key issues and formulated 12 strategies for sending inquiry requests as part of our study. From July 20 to August 1, 2024, we posed 11 surgery-related questions about SCR and requested the creation of preoperative patient education documents through the Poe website to 3 different LLMs: Claude-3-Opus, GPT-4-Turbo, and Gemini-1.5-Pro. These models collectively produced 33 responses and 3 preoperative patient education documents. From August 10 to August 17, 2024, three experienced sports medicine clinicians, who are not from the same institution, along with 5 patients who had undergone SCR surgery, evaluated the responses and documents provided by the LLMs.

Evaluations From the Subjective Perspective of Doctors

In total, 3 professional sports medicine doctors first evaluated the responses of 3 different LLMs to 11 inquiries. The evaluations focused on accuracy, completeness, logicality, potential risk, and overall rating. The results showed that Gemini's responses were significantly superior to Claude's in all evaluated categories including accuracy (mean 5.00, SD 0.00 vs mean 4.48, SD 0.83; $P<.001$), completeness (mean 4.88, SD 0.33 vs mean 4.39, SD 0.70; $P=.001$), logicality (mean 5.00, SD 0.00 vs mean 4.70, SD 0.59; $P<.01$) potential risk (mean 5.00, SD 0.00 vs mean 4.73, SD 0.57; $P<.01$), and overall rating (mean 9.88, SD 0.42 vs mean 9.03, SD 1.31; $P=.001$; Figures 2A and 2B). Compared to GPT, Gemini's responses were superior in all categories, with significant differences noted in completeness (mean 4.88, SD 0.33 vs mean 4.55, SD 0.67; $P=.02$), potential risk (mean 5.00, SD 0.00 vs mean 4.67, SD 0.82; $P=.01$), and overall rating (mean 9.88, SD 0.42 vs mean 9.24, SD 1.30; $P=.01$; Figures 2A and 2B. GPT's responses, when compared to Claude's, were superior in accuracy ($P=.03$), completeness ($P=.34$), logicality ($P=.11$), and overall rating ($P=.42$); however, Claude was rated higher in potential risk ($P=.85$; Figures 2A and 2B). Of these differences, only the accuracy presented a statistically significant difference (Figures 2A and 2B).

Figure 2. Quality evaluation results from doctors and patients for 11 questions generated by 3 large language models. (**A-B**) Evaluation from the doctor's perspective; (**C-D**) evaluation from the patient's perspective. n.s. not significant; *P<.05, **P<.01, ***P<.001.



In terms of the PEMAT scores for the preoperative patient education materials generated by each LLM, Gemini scored higher than GPT (mean 1.00, SD 0.00 vs mean 0.91, SD 0.09; P=.12), and GPT scored higher than Claude (mean 0.91, SD 0.09 vs mean 0.79, SD 0.10; P=.18), with only the difference between Gemini and Claude (mean 1.00, SD 0.00 vs mean 0.79, SD 0.10; P=.03) being statistically significant (Figure 3). Regarding the DISCERN scores, Claude achieved the highest overall score, followed by Gemini and then GPT, though these differences were not statistically significant (Table 2). In the item of the DISCERN which represents overall satisfaction (the 16th question presented in Table 2), Gemini scored the highest, while GPT and Claude scored the same, with no statistical significance in the differences. The consistency among the 3 evaluators was high, with no instances of "Poor agreement" or "Slight agreement" in their assessments (Multimedia Appendix 2).

Figure 3. PEMAT scoring percentage for the patient education document generated by three large language models. n.s.: not significant; *P<.05, **P<.01, ***P<.001.

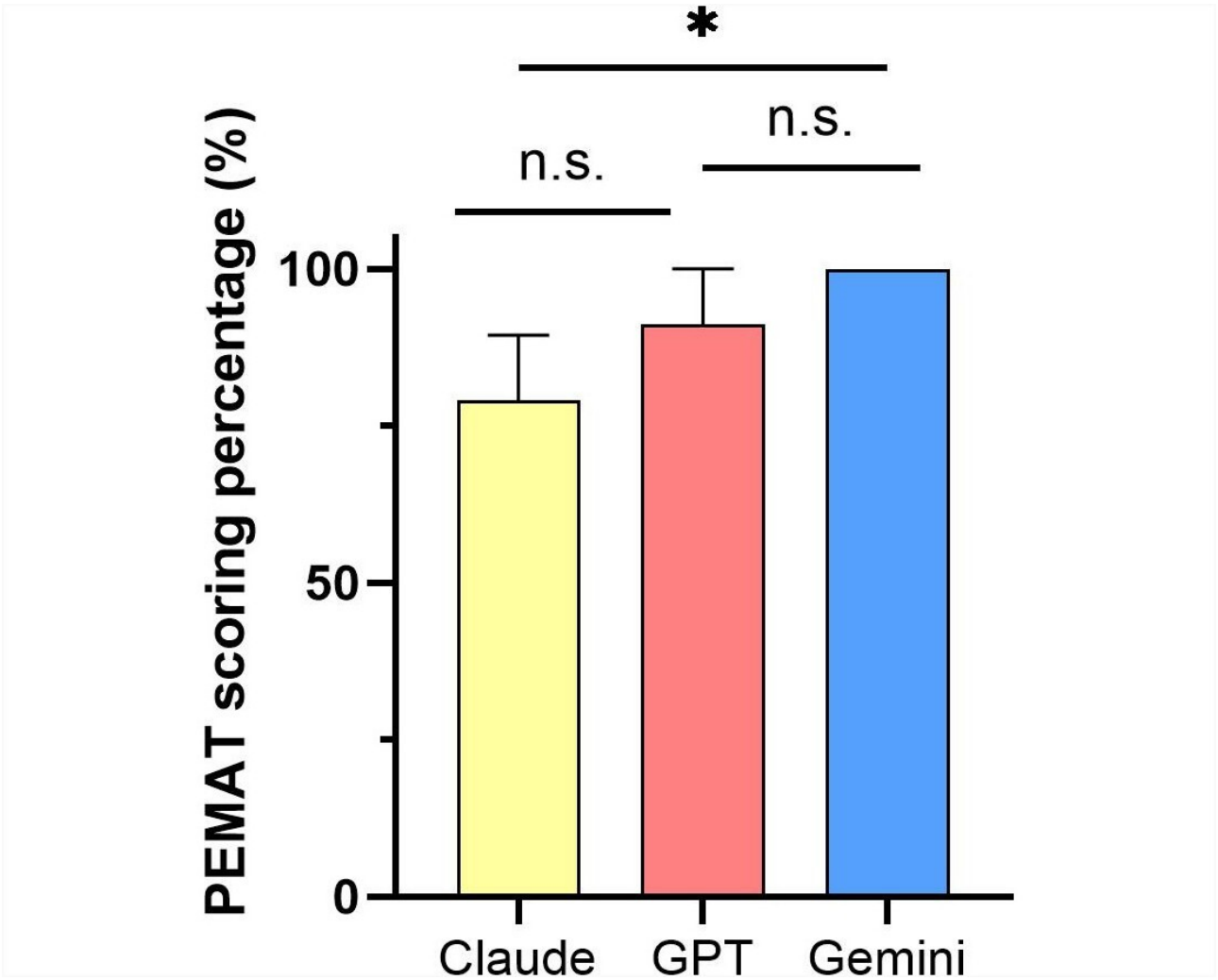


Table . Quality grades for section 2 of the DISCERN Tool.

Section 2. How good is the quality of information on treatment choices ?	Claude-3-Opus, Median (IQR)	GPT-4-Turbo, Median (IQR)	Gemini-1.5-Pro, Median (IQR)	Claude versus GPT, *P* value	Claude versus Gemini, *P* value	GPT versus Gemini, *P* value
Does it describe how each treatment works?	4 (3-4)	4 (3-4)	5 (4-5)	—[a]	.09	.09
Does it describe the benefits of each treatment?	4 (3-5)	4 (3-4)	1 (1-1)	.64	.04	.03
Does it describe the risks of each treatment?	4 (3-4)	3 (2-3)	5 (4-5)	.09	.09	.04
Does it describe what would happen if no treatment is used?	1 (1-1)	1 (1-1)	1 (1-1)	—	—	—
Does it describe how the treatment choices affect overall quality of life?	1 (1-1)	1 (1-1)	1 (1-1)	—	—	—
Is it clear that there may be more than one possible treatment choice?	1 (1-1)	1 (1-1)	1 (1-1)	—	—	—
Does it provide support for shared decision-making?	3 (3-4)	3 (2-3)	3 (2-3)	.32	.20	—
Based on the answers to all of the above questions, rate the overall quality of the publication as a source of information about treatment choices.	3 (3-4)	3 (3-4)	4 (3-4)	—	.46	.46

[a]Not applicable.

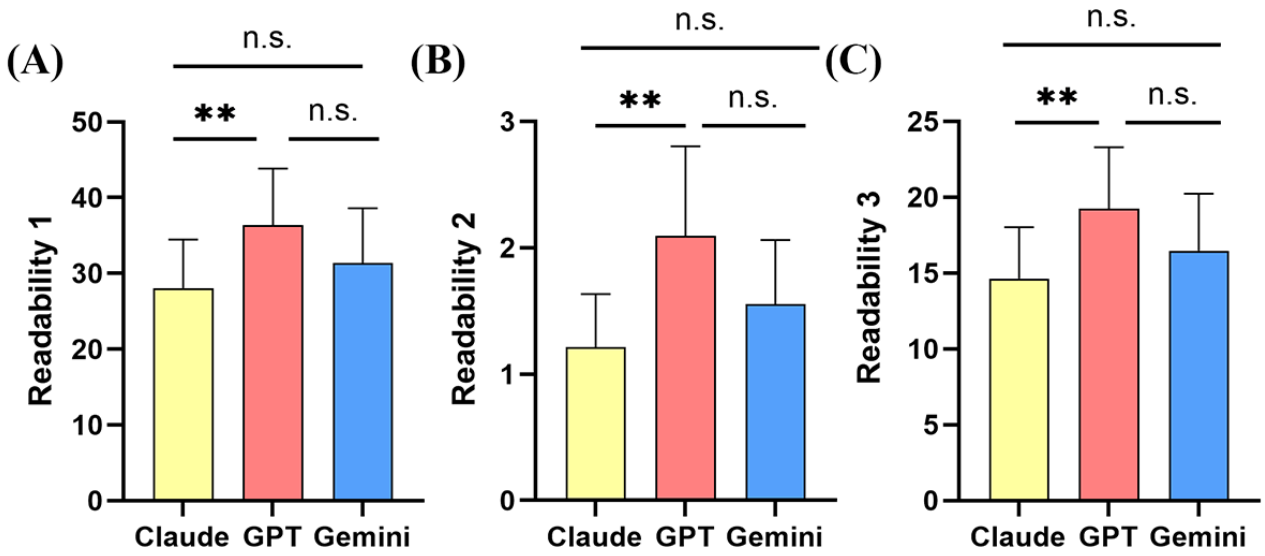
Evaluations From the Subjective Perspective of Patients

In the ratings provided by 5 follow-up patients for the preoperative patient education materials generated by the LLMs, Gemini scored higher than GPT and Claude across all parameters, including readability, educational quality, and overall rating (Figures 2C and 2D). Among these, the difference in educational quality between Gemini and Claude (mean 4.00, SD 0.00 vs mean 3.60, SD 0.55; *P*=.02) was statistically significant (Figures 2C and 2D). Furthermore, Gemini's advantage in overall satisfaction when compared to both Claude (mean 8.80, SD 0.45 vs mean 6.80, SD 1.10; *P*<.01) and GPT (mean 8.80, SD 0.45 vs mean 7.20, SD 0.84; *P*=.01) also showed statistical significance (Figures 2C and 2D). The consistency of all ratings given by the 5 follow-up patients was evaluated as "Fair agreement" (Multimedia Appendix 2).

Objective Evaluations of Readability

Based on the analysis methods of the context package, readability is assessed from 3 perspectives, namely readability 1, readability 2, and readability 3. Under these assessments, GPT's readability is higher than that of Gemini (readability 1: mean 36.38, SD 7.47 vs mean 31.39, SD 7.20, *P*=.18; readability 2: mean 2.09, SD 0.71 vs mean 1.55, SD 0.51, *P*=.09; readability 3: mean 19.24, SD 4.07 vs mean 16.47, SD 3.77, *P*=.17) and Claude (readability 1: mean 36.38, SD 7.47 vs mean 28.05, SD 6.43, *P*<.01; readability 2: mean 2.09, SD 0.71 vs mean 1.21, SD 0.42, *P*<.01; readability 3: mean 19.24, SD 4.07 vs mean 14.63, SD 3.40, *P*<.01), with the difference between GPT and Claude being statistically significant (Figure 4). Although Gemini's readability is higher than Claude's, the difference is not statistically significant (Figure 4). However, when readability was assessed using py-readability metrics, there was no statistical difference between the 3 LLM models (Multimedia Appendix 3).

Figure 4. Comparison of the results of text readability analysis from three analytical perspectives using the cntext package in R software. n.s.: not significant; *P<.05, **P<.01, ***P<.001.



Discussion

Principal Findings

The main findings of our study are as follows: (1) the three LLMs (Claude-3-Opus, GPT-4-Turbo, and Gemini-1.5-Pro) demonstrated good overall potential for application in patient education for SCR surgery. They were able to generate answers to 11 SCR-related questions and create standardized preoperative informed consent patient education documents. (2) In the subjective evaluations by professional sports medicine clinicians and patients who had undergone SCR surgery, Gemini slightly outperformed GPT and Claude in multiple dimensions, including accuracy, completeness, logic, potential risks, and overall satisfaction. (3) In this study, the 3 LLMs did not proactively provide evidence sources when answering questions and generating patient education documents. If LLMs are to be used to assist with patient education in clinical applications, it may be necessary to specifically require LLMs to cite information sources to enable doctors and patients to judge the authority and reliability of the content. (4) Although Gemini performed best in the ratings for SCR patient education-related tasks, considering the complexity and potential risks of LLMs in medical applications, clinicians still need to carefully review and make necessary corrections to the content generated by LLMs to ensure the professionalism and reasonableness of patient education materials. LLMs should be positioned as assistive tools rather than decision-making entities in clinical applications.

LLMs have proven to be reliable sources of information for orthopedic surgery-related questions, creating patient education documents that enhance the understanding of diagnostic and therapeutic processes for nonprofessionals and improve the readability of educational materials [28,37,38]. However, evaluating the quality of responses from LLMs is not straightforward. Researchers assessed ChatGPT 3.5's medical knowledge by using clinical standards and licensing examination questions to evaluate its theoretical understanding and practical application [39]. With the advent of ChatGPT 4.0 and the iterative upgrades of various LLMs from different companies, there has been a growing recognition and exploration of the expanded pretraining data and enhanced text processing capabilities of the latest LLM versions in different clinical scenarios [40,41]. Scholars have realized that the quality of LLM responses is influenced by multiple factors, including the amount of information in the query [42], the questioning strategy [43], and many unpredictable elements [44]. These unpredictable elements are evident when, under controlled conditions with all variables constant, the same question yields different answers and shows varying styles of text presentation. Consequently, while researchers have acknowledged the capabilities of LLMs in diagnosing, treating, and creating educational documents across disciplines, they continue to reject the idea of LLMs performing independent medical actions, affirming their role solely as an auxiliary tool in the hands of professionals [45,46].

This study aims to assess the feasibility of using three popular LLMs as auxiliary tools for sports medicine physicians during the informed consent process for patients undergoing SCR. In this study, physicians use LLMs primarily to assess the accuracy and comprehensiveness of the information and to clarify content. Unlike previous studies that evaluated answer readability solely through software analysis of word and sentence structure [4,6,47], this study also included follow-up visits with SCR patients post surgery, where patients subjectively assessed the readability and educational significance of the information. Patient ratings primarily focused on the presurgical educational materials generated by LLMs, excluding the evaluation of 11 specific questions, as the answers to these questions required physician assessment of accuracy and comprehensiveness and clarification before clinical use. Without this step by physicians, patients, who are not medical professionals, might not be able to accurately assess the details of the questions. Although all 3 models performed satisfactorily in evaluating "potential risks," this does not imply that patients can rely on LLMs as their sole source of medical advice. We believe that the SCR medical decision-making process, which does not involve extensive use of medications and auxiliary treatments pre- and post-surgery

and follows a "surgery-rehabilitation" model, does not necessitate the phase-wise, continuous assessments and patient education required for conditions like cancer.

Despite the potential benefits of using LLMs in patient education, several ethical and privacy issues need to be addressed before their widespread application. The accuracy and reliability of the information generated by LLMs are critical, especially in sensitive medical contexts. To enhance their accuracy, strategies such as retrieving pertinent information from credible, external data sources before generating text can be incorporated into subsequent versions of LLMs. And patient privacy is a fundamental concern when using LLMs in medical settings. LLMs may require access to patient data to generate personalized and relevant information. However, this access must be strictly regulated to prevent unauthorized use or disclosure of sensitive patient information.

In addition, our "Prompt Execution" phase revealed that without background information, LLMs occasionally misidentify SCR as a supraspinatus repair surgery under patch bridging, leading to content generation biases. We consider such biases to be system errors caused by human operational mistakes, which can be avoided by adjusting prompt strategies under the guidance of subject matter experts. Therefore, using LLMs for specialist information retrieval is not without its challenges, and we believe that merely relying on LLM-generated disclaimers like "I am not a medical professional; if you feel unwell, please seek medical attention immediately" at the end of responses is insufficient [28]. The mitigation of these errors can be facilitated through the use of techniques such as fine-tuning and retrieval-augmented generation. Fine-tuning entails training the LLM on a smaller, highly specialized dataset that has been meticulously curated to capture the intricate details of the medical domain and retrieval-augmented generation can address issues of hallucinations by first retrieving pertinent information from credible, external data sources before generating text. Incorporating these strategies into subsequent versions of LLMs has the potential to enhance their accuracy and reliability, particularly in sensitive applications such as patient education. A thorough examination would offer valuable insights into refining these models to deliver precise and trustworthy information within medical contexts.

Our study meanwhile discovers critical gaps in LLMs are used in medical settings, particularly in presurgical patient education. LLMs often do not provide sources for their information, and their responses can include inaccuracies or fabricated sources, known as "hallucinations" [48]. This issue is exacerbated when users do not specifically ask for sources, leading LLMs to sometimes provide outdated or irrelevant information [48,49]. Furthermore, the LLMs in the study failed to discuss alternative treatments, benefits, and risks associated with not undergoing specific surgeries like SCR. This omission is significant as discussing these elements is essential for informed medical decision-making and respects patient rights to understand all available options. Given these limitations, LLMs should not independently manage diagnosis or patient education. Instead, they should serve as supplementary tools, aiding health care professionals who can provide the necessary context, accuracy, and depth in patient interactions. This approach ensures that patient education remains thorough, accurate, and ethically conducted, aligning with medical standards and patient rights. This challenge can be tackled through the application of more advanced prompt engineering methodologies, the integration of contextual reasoning capabilities, and the implementation of step-by-step guidance mechanisms. By engaging in multiple iterative interactions with the model, it becomes possible to refine its responses and produce more comprehensive information, encompassing alternative treatment options, based on the specific inputs provided by the user. Such an approach would empower the LLM to deliver content that is more personalized, well-informed, and balanced. Moreover, the development of LLM-Agents offers a compelling solution to the limitations of LLMs in sensitive domains like medical decision-making. By integrating planning, memory, tool use, and agent or brain components, these agents can enhance their ability to provide accurate, verified information. This not only supports human expertise but also ensures that the information presented is transparent and evidence-backed. As research continues, the full potential of integrating citation capabilities within LLM-Agents should be explored to further improve their reliability and trustworthiness in high-stakes contexts.

With the evolution of internet technology, we have witnessed a transition from Web1.0 to Web2.0, and the ways we access information have dramatically changed—from relying on traditional media to accessing massive amounts of information anytime and anywhere via the internet, social media, and personal media platforms [50,51]. Particularly on social media and personal media platforms, we can find questions similar to our own and the corresponding responses [6,50,51]. However, the accuracy and comprehensiveness of information obtained in this manner can be uncertain [51]. Online responses vary greatly in quality, lacking systematic organization and authority, and the response time and outcomes of further inquiries are unpredictable. Studies have shown that answers from ChatGPT 3.5 are not only more comprehensive and empathetic than those from certified physicians on Reddit forums but, despite demonstrating high quality in assessing dementia care issues, they fall slightly short in predicting potential future problems [52,53]. When comparing responses from ChatGPT 4.0, 3.5, and those on Reddit, ChatGPT 4.0's responses significantly surpassed the others, reaching a new level of excellence [54]. In responding to patient inquiries, LLMs also perform more accurately than Google searches and are easier to read [27]. However, they also share a common drawback: the use of LLMs in medical consultations is best accompanied by professional medical personnel to "clarify" the responses [31]. Therefore, LLMs are not suitable for independently handling any part of the diagnostic or treatment process within the medical system, but they are better suited as tools to enhance the efficiency of professional medical personnel or as mediums for personalized patient communication and education [55,56].

As technology continues to advance, hospitals are consistently innovating in all aspects of clinical diagnosis and treatment to enhance diagnostic accuracy, treatment outcomes, and patient satisfaction, representing an unstoppable trend in health care innovation [57,58]. However, balancing standardized processes with personalized patient needs often presents a challenge [59].

LLMs present an opportunity to potentially maintain standardized quality in their responses while also accommodating personalized requests. LLMs, encompassing both free and paid versions, are generally accessible to the public as open platforms [60]. Although current research does not support its use in guiding clinical decisions [61], using ChatGPT in doctor-patient communication benefits both doctors and patients [7]. Doctors can interpret and supplement ChatGPT's responses based on their clinical experience, offering more personalized consultations to patients [31]. In addition, patients reduce their need to search for information on the internet, and their trust in physicians may be enhanced with the objective evidence provided by AI. Under the joint oversight of doctors and patients, the advantages of artificial intelligence can be fully used [62]. Nevertheless, the widespread adoption and application of LLMs still face technical and policy limitations. Technical limitations include differences in handling inputs in various languages [63], performance discrepancies between proprietary and open-source models [64], and the occurrence of "hallucinations" when faced with biased questions [65]. Since commonly used LLMs like GPT, Gemini, and Claude are proprietary, and these models are trained with significantly more data than open-source models, we can only continue to explore ways to avoid "hallucinations" instead of fixing the root cause of such issues [66,67]. In addition, policy restrictions cannot be ignored [68]. Health systems and hospitals need to develop detailed policies to regulate the clinical auxiliary use of LLMs, including ensuring patient informed consent, standardized user training, and the preservation of usage records [7]. Sound policies are essential to ensure the appropriate and efficient use of tools [65,68]. Through these measures, the safety of LLM applications in the medical field can be effectively enhanced, protecting patient rights while improving the efficiency and quality of doctor-patient communication [47,69].

Limitations

This study has several limitations. First, both the linguistic input and the analyzed responses were in Chinese. On one hand, this choice was made to facilitate assessments by Chinese-speaking clinical experts and patients during follow-ups. On the other hand, input in different languages could introduce potential errors and biases. Second, this research only explores the feasibility of using LLMs to generate content related to SCR for patient education. The variability in surgical procedures and specialties could pose distinct challenges in patient education, which means the conclusions drawn from this study cannot be simply generalized to other disciplines. Finally, during the "Prompts Development" phase, it was found that without additional background information, SCRs are prone to be misidentified by LLMs as bridge suture repairs of the supraspinatus muscle. However, since all 3 models used were proprietary, we opted for a "Background+ Question" approach to mitigate this systematic error, without being able to investigate the reasons behind such occurrences.

Conclusions

Claude-3-Opus, GPT-4-Turbo, and Gemini-1.5-Pro effectively addressed patient queries and generated readable presurgical education materials. However, they lacked citations and failed to explore alternative treatments, benefits, and potential risks of forgoing SCR surgery. While these LLMs can serve as valuable aids for physicians, they should not be used as standalone tools for patient education without expert oversight to ensure comprehensive and accurate information is provided.

Acknowledgments

We would like to express our deepest gratitude to all the experts and patients who have contributed to this research.

Data Availability

All data included in this study are available upon request by contact with the corresponding author.

The subjects of this study are LLMs (large language models). Besides being used as operational models, LLMs also serve as tools for translating Chinese content into English, as detailed in Multimedia Appendix 1. The specific types of models used, the websites they are accessed through, and their methods of use are all mentioned in the relevant sections. Beyond these functions, LLMs do not influence the generation of the article's content in any other way.

Authors' Contributions

Conceptualization: WY Gan, H Li, JF Ouyang
Methodology: WY Gan, H Li, JF Ouyang
Supervision: XF Zheng
Visualization: YK Liu
Writing—original draft: WY Gan, H Li, JF Ouyang, YK Liu
Writing—reviewing and editing: WY Gan, H Li, JF Ouyang, YK Liu, ZW Xue, M Wang, HB He, B Song, XF Zheng

Conflicts of Interest

None declared.

Multimedia Appendix 1

All Questions and Answers for Claude-3-Opus, GPT-4-Turbo, and Gemini-1.5-Pro (Use GPT-4-Turbo for Chinese to English translation).
[DOCX File, 71 KB - periop_v8i1e70047_app1.docx]

Multimedia Appendix 2
Table S1: Consistent evaluation of Fleiss kappa among raters.
[DOCX File, 14 KB - periop_v8i1e70047_app2.docx]

Multimedia Appendix 3
Comaprison of readability by py-readability-metrics.
[DOCX File, 19 KB - periop_v8i1e70047_app3.docx]

References

1. Flaharty KA, Hu P, Hanchard SL, et al. Evaluating large language models on medical, lay-language, and self-reported descriptions of genetic conditions. Am J Hum Genet 2024 Sep 5;111(9):1819-1833. [doi: 10.1016/j.ajhg.2024.07.011] [Medline: 39146935]
2. Rengers TA, Thiels CA, Salehinejad H. Academic Surgery in the Era of Large Language Models: A Review. JAMA Surg 2024 Apr 1;159(4):445-450. [doi: 10.1001/jamasurg.2023.6496] [Medline: 38353991]
3. Chow R, Hasan S, Zheng A, et al. The Accuracy of Artificial Intelligence ChatGPT in Oncology Examination Questions. J Am Coll Radiol 2024 Nov;21(11):1800-1804. [doi: 10.1016/j.jacr.2024.07.011] [Medline: 39098369]
4. Eng E, Mowers C, Sachdev D, et al. Chat Generative Pre-Trained Transformer (ChatGPT) – 3.5 Responses Require Advanced Readability for the General Population and May Not Effectively Supplement Patient-Related Information Provided by the Treating Surgeon Regarding Common Questions About Rotator Cuff Repair. Arthroscopy: The Journal of Arthroscopic & Related Surgery 2025 Jan;41(1):42-52. [doi: 10.1016/j.arthro.2024.05.009]
5. Mika AP, Martin JR, Engstrom SM, Polkowski GG, Wilson JM. Assessing ChatGPT Responses to Common Patient Questions Regarding Total Hip Arthroplasty. Journal of Bone and Joint Surgery 2023;105(19):1519-1526. [doi: 10.2106/JBJS.23.00209]
6. Pan A, Musheyev D, Bockelman D, Loeb S, Kabarriti AE. Assessment of Artificial Intelligence Chatbot Responses to Top Searched Queries About Cancer. JAMA Oncol 2023 Oct 1;9(10):1437-1440. [doi: 10.1001/jamaoncol.2023.2947] [Medline: 37615960]
7. Xue Z, Zhang Y, Gan W, Wang H, She G, Zheng X. Quality and Dependability of ChatGPT and DingXiangYuan Forums for Remote Orthopedic Consultations: Comparative Analysis. J Med Internet Res 2024 Mar 14;26:e50882. [doi: 10.2196/50882] [Medline: 38483451]
8. Gertz RJ, Dratsch T, Bunck AC, et al. Potential of GPT-4 for detecting errors in radiology reports: Implications for reporting accuracy. Radiology 2024 Apr;311(1):e232714. [doi: 10.1148/radiol.232714] [Medline: 38625012]
9. Maida M, Ramai D, Mori Y, et al. The role of generative language systems in increasing patient awareness of colon cancer screening. Endoscopy 2025 Mar;57(3):262-268. [doi: 10.1055/a-2388-6084] [Medline: 39142348]
10. Ebel S, Ehrengut C, Denecke T, Gößmann H, Beeskow AB. GPT-4o's competency in answering the simulated written European Board of Interventional Radiology exam compared to a medical student and experts in Germany and its ability to generate exam items on interventional radiology: a descriptive study. J Educ Eval Health Prof 2024;21:21. [doi: 10.3352/jeehp.2024.21.21] [Medline: 39161266]
11. Gan W, Ouyang J, Li H, et al. Integrating ChatGPT in orthopedic education for medical undergraduates: Randomized controlled trial. J Med Internet Res 2024 Aug 20;26:e57037. [doi: 10.2196/57037] [Medline: 39163598]
12. Mihata T, McGarry MH, Pirolo JM, Kinoshita M, Lee TQ. Superior capsule reconstruction to restore superior stability in irreparable rotator cuff tears: a biomechanical cadaveric study. Am J Sports Med 2012 Oct;40(10):2248-2255. [doi: 10.1177/0363546512456195] [Medline: 22886689]
13. E. Cline K, Tibone JE, Ihn H, et al. Superior Capsule Reconstruction Using Fascia Lata Allograft Compared With Double- and Single-Layer Dermal Allograft: A Biomechanical Study. Arthroscopy: The Journal of Arthroscopic & Related Surgery 2021 Apr;37(4):1117-1125. [doi: 10.1016/j.arthro.2020.11.054]
14. Mihata T, Lee TQ, Hasegawa A, et al. Arthroscopic superior capsule reconstruction for irreparable rotator cuff tears: Comparison of clinical outcomes with and without subscapularis tear. Am J Sports Med 2020 Dec;48(14):3429-3438. [doi: 10.1177/0363546520965993] [Medline: 33104385]
15. Claro R, Fonte H. Superior capsular reconstruction: current evidence and limits. EFORT Open Rev 2023 May 9;8(5):340-350. [doi: 10.1530/EOR-23-0027] [Medline: 37158430]
16. Mihata T, Lee TQ, Watanabe C, et al. Clinical results of arthroscopic superior capsule reconstruction for irreparable rotator cuff tears. Arthroscopy: The Journal of Arthroscopic & Related Surgery 2013 Mar;29(3):459-470. [doi: 10.1016/j.arthro.2012.10.022]

17. Hirahara AM, Andersen WJ, Panero AJ. Superior capsular reconstruction: Clinical outcomes after minimum 2-year follow-up. Am J Orthop (Belle Mead NJ) 2017;46(6):266-278. [Medline: 29309442]
18. Snyder SJ, Arnoczky SP, Bond JL, Dopirak R. Histologic evaluation of a biopsy specimen obtained 3 months after rotator cuff augmentation with GraftJacket Matrix. Arthroscopy: The Journal of Arthroscopic & Related Surgery 2009 Mar;25(3):329-333. [doi: 10.1016/j.arthro.2008.05.023]
19. Edwards PK, Mears SC, Lowry Barnes C. Preoperative education for hip and knee replacement: Never stop learning. Curr Rev Musculoskelet Med 2017 Sep;10(3):356-364. [doi: 10.1007/s12178-017-9417-4] [Medline: 28647838]
20. Alattas SA, Smith T, Bhatti M, Wilson-Nunn D, Donell S. Greater pre-operative anxiety, pain and poorer function predict a worse outcome of a total knee arthroplasty. Knee Surg Sports Traumatol Arthrosc 2017 Nov;25(11):3403-3410. [doi: 10.1007/s00167-016-4314-8] [Medline: 27734110]
21. Krebs ED, Hoang SC. Informed consent and shared decision making in the perioperative environment. Clin Colon Rectal Surg 2023 May;36(03):223-228. [doi: 10.1055/s-0043-1761158]
22. Noble PC, Fuller-Lafreniere S, Meftah M, Dwyer MK. Challenges in outcome measurement: Discrepancies between patient and provider definitions of success. Clin Orthop Relat Res 2013;471(11):3437-3445. [doi: 10.1007/s11999-013-3198-x]
23. Villanueva C, Talwar A, Doyle M. Improving informed consent in cardiac surgery by enhancing preoperative education. Patient Educ Couns 2018 Dec;101(12):2047-2053. [doi: 10.1016/j.pec.2018.06.008] [Medline: 29937111]
24. Bollschweiler E, Apitzsch J, Obliers R, et al. Improving informed consent of surgical patients using a multimedia-based program? Results of a prospective randomized multicenter study of patients before cholecystectomy. Ann Surg 2008 Aug;248(2):205-211. [doi: 10.1097/SLA.0b013e318180a3a7] [Medline: 18650629]
25. Sceats LA, Morris AM, Narayan RR, Mezynski A, Woo RK, Yang GP. Lost in translation: Informed consent in the medical mission setting. Surgery 2019 Feb;165(2):438-443. [doi: 10.1016/j.surg.2018.06.010] [Medline: 30061041]
26. Neubauer PD, Tabaee A, Schwam ZG, Francis FK, Manes RP. Patient knowledge and expectations in endoscopic sinus surgery. Int Forum Allergy Rhinol 2016 Sep;6(9):921-925. [doi: 10.1002/alr.21763] [Medline: 27028979]
27. Hristidis V, Ruggiano N, Brown EL, Ganta SRR, Stewart S. ChatGPT vs Google for queries related to dementia and other cognitive decline: Comparison of results. J Med Internet Res 2023 Jul 25;25:e48966. [doi: 10.2196/48966] [Medline: 37490317]
28. Oeding JF, Lu AZ, Mazzucco M, et al. ChatGPT-4 Performs clinical information retrieval tasks using consistently more trustworthy resources than does google search for queries concerning the Latarjet procedure. Arthroscopy: The Journal of Arthroscopic & Related Surgery 2025 Mar;41(3):588-597. [doi: 10.1016/j.arthro.2024.05.025]
29. Nicikowski J, Szczepański M, Miedziaszczyk M, Kudliński B. The potential of ChatGPT in medicine: an example analysis of nephrology specialty exams in Poland. Clin Kidney J 2024 Aug;17(8):sfae193. [doi: 10.1093/ckj/sfae193] [Medline: 39099569]
30. Bernstein IA, Zhang YV, Govil D, et al. Comparison of ophthalmologist and large language model chatbot responses to online patient eye care questions. JAMA Netw Open 2023 Aug 1;6(8):e2330320. [doi: 10.1001/jamanetworkopen.2023.30320] [Medline: 37606922]
31. Li LT, Sinkler MA, Adelstein JM, Voos JE, Calcei JG. Chatgpt responses to common questions about anterior cruciate ligament reconstruction are frequently satisfactory. Arthroscopy: The Journal of Arthroscopic & Related Surgery 2024 Jul;40(7):2058-2066. [doi: 10.1016/j.arthro.2023.12.009]
32. Nwachukwu BU, Varady NH, Allen AA, et al. Currently available large language models do not provide musculoskeletal treatment recommendations that are concordant with evidence-based clinical practice guidelines. Arthroscopy: The Journal of Arthroscopic & Related Surgery 2025 Feb;41(2):263-275. [doi: 10.1016/j.arthro.2024.07.040]
33. Chen L, Zaharia M, Zou J. How is chatgpt's behavior changing over time? . Preprint posted online on Jul 1, 2023 URL: https://ui.adsabs.harvard.edu/abs/2023arXiv230709009C [accessed 2025-06-06]
34. Menz BD, Kuderer NM, Bacchi S, et al. Current safeguards, risk mitigation, and transparency measures of large language models against the generation of health disinformation: repeated cross sectional analysis. BMJ 2024 Mar 20;384:e078538. [doi: 10.1136/bmj-2023-078538] [Medline: 38508682]
35. Yalamanchili A, Sengupta B, Song J, et al. Quality of large language model responses to radiation oncology patient care questions. JAMA Netw Open 2024 Apr 1;7(4):e244630. [doi: 10.1001/jamanetworkopen.2024.4630] [Medline: 38564215]
36. Li HQ, Yang Y, Xue FW, Liu ZY. Annual report readability and trade credit financing: Evidence from China. Research in International Business and Finance 2024 Apr;69:102220. [doi: 10.1016/j.ribaf.2024.102220]
37. Draschl A, Hauer G, Fischerauer SF, et al. Are chatgpt's free-text responses on periprosthetic joint infections of the hip and knee reliable and useful? J Clin Med 2023 Oct 20;12(20):6655. [doi: 10.3390/jcm12206655] [Medline: 37892793]
38. Kaarre J, Feldt R, Keeling LE, et al. Exploring the potential of ChatGPT as a supplementary tool for providing orthopaedic information. Knee surg sports traumatol arthrosc 2023 Nov;31(11):5190-5198. [doi: 10.1007/s00167-023-07529-2]
39. Sumbal A, Sumbal R, Amir A. Can ChatGPT-3.5 pass a medical exam? A systematic review of ChatGPT's performance in academic testing. J Med Educ Curric Dev 2024;11(23821205241238641):23821205241238641. [doi: 10.1177/23821205241238641] [Medline: 38487300]

40. Deng L, Wang T, et al. Evaluation of large language models in breast cancer clinical scenarios: a comparative analysis based on ChatGPT-3.5, ChatGPT-4.0, and Claude2. Int J Surg 2024 Jan;110(4):1941-1950. [doi: 10.1097/JS9.0000000000001066]
41. Jarry Trujillo C, Vela Ulloa J, Escalona Vivas G, et al. Surgeons vs ChatGPT: Assessment and feedback performance based on real surgical scenarios. J Surg Educ 2024 Jul;81(7):960-966. [doi: 10.1016/j.jsurg.2024.03.012] [Medline: 38749814]
42. Zhu L, Mou W, Lai Y, et al. Step into the era of large multimodal models: a pilot study on ChatGPT-4V(ision)'s ability to interpret radiological images. Int J Surg 2024 Jul 1;110(7):4096-4102. [doi: 10.1097/JS9.0000000000001359] [Medline: 38498394]
43. Lim ZW, Pushpanathan K, Yew SME, et al. Benchmarking large language models' performances for myopia care: a comparative analysis of ChatGPT-3.5, ChatGPT-4.0, and Google Bard. EBioMedicine 2023 Sep;95(104770):104770. [doi: 10.1016/j.ebiom.2023.104770] [Medline: 37625267]
44. Chervenak J, Lieman H, Blanco-Breindel M, Jindal S. The promise and peril of using a large language model to obtain clinical information: ChatGPT performs strongly as a fertility counseling tool with limitations. Fertil Steril 2023 Sep;120(3 Pt 2):575-583. [doi: 10.1016/j.fertnstert.2023.05.151] [Medline: 37217092]
45. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. Nat Med 2023 Aug;29(8):1930-1940. [doi: 10.1038/s41591-023-02448-8] [Medline: 37460753]
46. Tan S, Xin X, Wu D. ChatGPT in medicine: prospects and challenges: a review article. Int J Surg 2024 Jun 1;110(6):3701-3706. [doi: 10.1097/JS9.0000000000001312] [Medline: 38502861]
47. Haver HL, Gupta AK, Ambinder EB, et al. Evaluating the use of ChatGPT to accurately simplify patient-centered information about breast cancer prevention and screening. Radiol Imaging Cancer 2024 Mar;6(2):e230086. [doi: 10.1148/rycan.230086] [Medline: 38305716]
48. Chelli M, Descamps J, Lavoué V, et al. Hallucination rates and reference accuracy of ChatGPT and Bard for systematic reviews: Comparative analysis. J Med Internet Res 2024 May 22;26:e53164. [doi: 10.2196/53164] [Medline: 38776130]
49. Burnette H, Pabani A, von Itzstein MS, et al. Use of artificial intelligence chatbots in clinical management of immune-related adverse events. J Immunother Cancer 2024 May 30;12(5):38816231. [doi: 10.1136/jitc-2023-008599] [Medline: 38816231]
50. Terrasse M, Gorin M, Sisti D. Social media, e-health, and medical ethics. Hastings Cent Rep 2019 Jan;49(1):24-33. [doi: 10.1002/hast.975] [Medline: 30790306]
51. Ho A, McGrath C, Mattheos N. Social media patient testimonials in implant dentistry: information or misinformation? Clin Oral Implants Res 2017 Jul;28(7):791-800. [doi: 10.1111/clr.12883] [Medline: 27279455]
52. Ayers JW, Poliak A, Dredze M, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Intern Med 2023 Jun 1;183(6):589-596. [doi: 10.1001/jamainternmed.2023.1838] [Medline: 37115527]
53. Aguirre A, Hilsabeck R, Smith T, et al. Assessing the quality of chatgpt responses to dementia caregivers' questions: Qualitative analysis. JMIR Aging 2024 May 6;7:e53019. [doi: 10.2196/53019] [Medline: 38722219]
54. Girton MR, Greene DN, Messerlian G, Keren DF, Yu M. ChatGPT vs medical professional: Analyzing responses to laboratory medicine questions on social media. Clin Chem 2024 Sep 3;70(9):1122-1139. [doi: 10.1093/clinchem/hvae093] [Medline: 39013110]
55. La Bella S, Attanasi M, Porreca A, et al. Reliability of a generative artificial intelligence tool for pediatric familial Mediterranean fever: insights from a multicentre expert survey. Pediatr Rheumatol Online J 2024 Aug 23;22(1):78. [doi: 10.1186/s12969-024-01011-0] [Medline: 39180115]
56. Cavnar Helvaci B, Hepsen S, Candemir B, et al. Assessing the accuracy and reliability of ChatGPT's medical responses about thyroid cancer. Int J Med Inform 2024 Nov;191(105593):105593. [doi: 10.1016/j.ijmedinf.2024.105593] [Medline: 39151245]
57. Pallett AC, Nguyen BT, Klein NM, Phippen N, Miller CR, Barnett JC. A randomized controlled trial to determine whether A video presentation improves informed consent for hysterectomy. Am J Obstet Gynecol 2018 Sep;219(3):277. [doi: 10.1016/j.ajog.2018.06.016] [Medline: 29959929]
58. Zhang MH, Haq ZU, Braithwaite EM, Simon NC, Riaz KM. A randomized, controlled trial of video supplementation on the cataract surgery informed consent process. Graefes Arch Clin Exp Ophthalmol 2019 Aug;257(8):1719-1728. [doi: 10.1007/s00417-019-04372-5] [Medline: 31144057]
59. McCollough CH. Standardization versus individualization: how each contributes to managing dose in computed tomography. Health Phys 2013 Nov;105(5):445-453. [doi: 10.1097/HP.0b013e31829db936] [Medline: 24077044]
60. Vaid A, Duong SQ, Lampert J, et al. Local large language models for privacy-preserving accelerated review of historic echocardiogram reports. J Am Med Inform Assoc 2024 Sep 1;31(9):2097-2102. [doi: 10.1093/jamia/ocae085] [Medline: 38687616]
61. Balla Y, Tirunagari S, Windridge D. Machine learning in pediatrics: Evaluating challenges, opportunities, and explainability. Indian Pediatr 2023 May 14. [Medline: 37179470]
62. Yeo YH, Samaan JS, Ng WH, et al. Assessing the performance of ChatGPT in answering questions regarding cirrhosis and hepatocellular carcinoma. Clin Mol Hepatol 2023 Jul;29(3):721-732. [doi: 10.3350/cmh.2023.0089] [Medline: 36946005]

63. Shao CY, Li H, Liu XL, et al. Appropriateness and comprehensiveness of using ChatGPT for perioperative patient education in thoracic surgery in different language contexts: Survey study. Interact J Med Res 2023 Aug 14;12:e46900. [doi: 10.2196/46900] [Medline: 37578819]
64. Sandmann S, Riepenhausen S, Plagwitz L, Varghese J. Systematic analysis of ChatGPT, Google search and Llama 2 for clinical decision support tasks. Nat Commun 2024 Mar 6;15(1):2050. [doi: 10.1038/s41467-024-46411-8] [Medline: 38448475]
65. Rao A, Pang M, Kim J, et al. Assessing the utility of ChatGPT throughout the entire clinical workflow: Development and usability study. J Med Internet Res 2023 Aug 22;25:e48659. [doi: 10.2196/48659] [Medline: 37606976]
66. Masters K. Medical Teacher 's first ChatGPT's referencing hallucinations: Lessons for editors, reviewers, and teachers. Med Teach 2023 Jul 3;45(7):673-675. [doi: 10.1080/0142159X.2023.2208731]
67. Hatem R, Simmons B, Thornton JE. A call to address AI "Hallucinations" and how healthcare professionals can mitigate their risks. Cureus 2023 Sep;15(9):37809168. [doi: 10.7759/cureus.44720]
68. Bukar UA, Sayeed MS, Razak SFA, Yogarayan S, Amodu OA. An integrative decision-making framework to guide policies on regulating ChatGPT usage. PeerJ Comput Sci 2024;10:e1845. [doi: 10.7717/peerj-cs.1845] [Medline: 38440047]
69. Platt J, Nong P, Smiddy R, et al. Public comfort with the use of ChatGPT and expectations for healthcare. J Am Med Inform Assoc 2024 Sep 1;31(9):1976-1982. [doi: 10.1093/jamia/ocae164]

Abbreviations

LLM: large language model
PEMAT: Patient Education Materials Assessment Tool
SCR: superior capsular reconstruction

Edited by N Rohatgi; submitted 13.12.24; peer-reviewed by FT Johora Faria, M Ma; revised version received 04.04.25; accepted 08.04.25; published 12.06.25.

Please cite as:
Liu Y, Li H, Ouyang J, Xue Z, Wang M, He H, Song B, Zheng X, Gan W
Evaluating Large Language Models for Preoperative Patient Education in Superior Capsular Reconstruction: Comparative Study of Claude, GPT, and Gemini
JMIR Perioper Med 2025;8:e70047
URL: https://periop.jmir.org/2025/1/e70047
doi:10.2196/70047

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