Towards Prescriptive Analytics Systems in Healthcare Delivery: AI-Transformation to Improve High Volume Operating Rooms Throughput

by

Farid Al Zoubi

Thesis submitted to the University of Ottawa
in partial Fulfillment of the requirements for the

Doctorate in Philosophy
Electrical and Computer Engineering

School of Electrical Engineering and Computer Science
Faculty of Engineering
University of Ottawa

Abstract

The increasing demand for healthcare services, coupled with the challenges of managing budgets and navigating complex regulations, has underscored the need for sustainable and efficient healthcare delivery. In response to this pressing issue, this thesis aims to optimize hospital efficiency using Artificial Intelligence (AI) techniques. The focus extends beyond improving surgical intraoperative time to encompass preoperative and postoperative periods as well.

The research presents a novel Prescriptive Analytics System (PAS) designed to enhance the Surgical Success Rate (SSR) in surgeries and specifically in high volume arthroplasty. The SSR is a critical metric that reflects the successful completion of 4-surgeries during an 8-hour timeframe. By leveraging AI, the developed PAS has the potential to significantly improve the SSR from its current rate of 39% at The Ottawa Hospital to a remarkable 100%.

The research is structured around five peer-reviewed journal papers, each addressing a specific aspect of the optimization of surgical efficiency. The first paper employs descriptive analytics to examine the factors influencing delays and overtime pay during surgeries. By identifying and analyzing these factors, insights are gained into the underlying causes of surgery inefficiencies.

The second paper proposes three frameworks aimed at improving Operating Room (OR) throughput. These frameworks provide structured guidelines and strategies to enhance the overall efficiency of surgeries, encompassing preoperative, intraoperative, and postoperative stages. By streamlining the workflow and minimizing bottlenecks, the proposed frameworks have the potential to significantly optimize surgical operations.

The third paper outlines a set of actions required to transform a selected predictive system into a prescriptive one. By integrating AI algorithms with decision support mechanisms, the system can offer actionable recommendations to surgeons during surgeries. This transformative step holds tremendous potential in enhancing surgical outcomes while reducing time.

The fourth paper introduces a benchmarking and monitoring system for the selected framework that predicts SSR. Leveraging historical data, this system utilizes supervised machine learning algorithms to forecast the likelihood of successful outcomes based on various surgical team and procedural parameters. By providing real-time monitoring and predictive insights, surgeons can proactively address potential risks and improve decision-making during surgeries.

Lastly, an application paper demonstrates the practical implementation of the prescriptive analytics system. The case study highlights how the system optimizes the allocation of resources and enables the scheduling of additional surgeries on days with a high predicted SSR. By leveraging the system's capabilities, hospitals can maximize their surgical capacity and improve overall patient care.

Acknowledgements

First and foremost, to God for granting me the health and patience to complete this work. It is through His fixed support and guidance that I have been able to overcome challenges and persevere. I am humbled by His grace, which has been my constant source of inspiration and resilience. May His blessings continue to inspire and guide me in all my endeavors.

My deepest gratitude to my esteemed professor, Pascal Fallavollita. Throughout my journey, his collaboration, exceptional knowledge, and understanding of the research problem have been invaluable. I am immensely grateful for his continuous guidance and mentorship, which has significantly contributed to the successful completion of this thesis.

I would like to thank our unwavering clinical collaborator, Dr. Paul Beaule, Chief of Staff for Hawkesbury General Hospital and Director of Research and Innovation for The Ottawa Hospital’s Division of Orthopedics. His clinical leadership and willingness to change current practice within the Division of Orthopedics has enhanced this thesis work exponentially.

I would also like to take a moment to remember and pay tribute to my late father, my late mother, and my late brothers who passed away before witnessing the completion of this significant milestone in my academic career. My mother, in particular, was exceptionally keen on my academic pursuits and consistently motivated me to conquer challenges. Her unconditional love and unwavering belief in my potential propelled me forward, even during the most difficult times. I carry her memory with me, cherishing the impact she had on shaping my determination and the ability to adapt.

To my dearest sister Karima, my nieces LilyAnn and Zaina, and my nephews Ibrahim, Yousef and Zaid, you have consistently brought joy to my heart, even during the most challenging times I've faced.

In addition to my professor and my departed family members, I would like to extend my appreciation to all my friends, colleagues, and classmates who stood by me during these endeavors. Your encouragement, discussions, and companionship played a vital role in shaping my ideas and refining my research.

To all those who have contributed to my Ph.D. journey, your presence and assistance have made this achievement possible. Thank you from the bottom of my heart.

Publications

Table 1: Thesis contribution in the field and publications

No	Title	Short name	Authorship	Type	Year	Journal/Conference	Status
1	*Factors Influencing Delays and Overtime During Surgery: A Descriptive Analytics for High Volume Arthroplasty Procedures*	J1	1st Author	Journal	2024	*Frontiers in Surgery*	**Published**
2	*Frameworks for AI-based Surgical Transformation (FAST)*	J2	1st Author	Journal	2023	*Frontiers in AI*	Under review
3	*Use of multidisciplinary positive deviance seminars to improve efficiency in a high-volume arthroplasty practice: a pilot study*	J3	1st Co-Author	Journal	2022	*Canadian Journal of Surgery*	**Published**
4	*Artificial-Intelligence driven prescriptive model to optimize team efficiency in a high-volume primary arthroplasty practice*	J4	1st Author	Journal	2023	*International Orthopaedics*	**Published**
5	*Leveraging Machine Learning and Prescriptive Analytics to Improve Operating Room Throughput*	J5	1st Author	Journal	2023	*Frontiers in Digital Health*	**Published**
6	*First Deployment of Artificial Intelligence Recommendations in Orthopedic Surgery*	J6	1st Author	Journal	2024	*Frontiers in Artificial Intelligence Medicine and Public Health*	**Published**
7	*Non-Clinical Data Science*	J7	1st Author	Journal	2024		To be submitted
8	*Perspectives: Changing philosophy in the workplace*	J8	1st Author	Journal	2023		To be submitted
9	*Artificial-intelligence Driven Benchmarks to Optimize Team Efficiency in Arthroplasty*	A1	2nd Author	Conference abstract	2022	*Orthopedic Research Society*	**Published**
10	*Use of Artificial Intelligence-Machine Learning to improve team efficiency in the arthroplasty operating room: a benchmark-based model*	A2	2nd Author	Conference abstract	2021	*Canadian Orthopedic Association Annual Meeting*	**Published**

Contents

List of Abbreviations

AACD	**Association of Anesthesia Clinical Directors**
AFT	**Anaesthesia Finish Time**
AI	**Artificial Intelligence**
APT	**Anaesthesia Preparation Time**
APT In room	**APT finished and patient in OR**
ASA	**American Society of Anesthesiologists**
AUC	**Area Under the ROC Curve**
CHT	**Canada Health Transfer**
CV	**Cross Validation**
DA	**Descriptive Analytics**
DNN-ANN	**Deep Neural Network-Artificial Neural Networks with multiple hidden layer**
DT	**Decision Tree**
EECS	**Electrical Engineering and Computer Science**
EQ-5D	**EuroQuol five-dimension**
FAST	**Frameworks for AI-based Surgical Transformation**
HR	**Human Resource**
HRA	**Hip Resurfacing Arthroplasty**
LG	**Logistic Regression**
ML	**Machine Learning**
OR	**Operating Room**
PA	**Prescriptive Analytics**
PACU	**Post Anesthesia Care Unit**
PAS	**Prescriptive Analytics System**
Procedure	**Actual Surgical Time**
PSF	**Patient Scheduling Framework**
RF	**Random Forest**
RN	**Registered Nurse**
ROC	**Receiver Operating Characteristic**
SDCH	**Surgical Day Care Unit**
SFT	**Surgery finish time**
SIMS	**Surgical Information Management System**
SPT	**Surgical Preparation Time**
SSR	**Surgical Success Rate**
STSF	**Surgical Team Scheduling Framework**
SVM	**Support Vector Machine**
THA	**Total Hip Arthroplasty**
THO	**The Ottawa Hospital**
TKA	**Total Knee Arthroplasty**
Turnover	**First patient exits to subsequent patient in room**
UKA	**Unicompartmental Knee Arthroplasty**
WTMF	**Workflow/Time Monitoring Framework**
XGBoost	**Extreme Gradient Boosting: Gradient-Boosted Decision Tree (GBDT)**

Chapter 1: Introduction

In this chapter, a fundamental overview of the healthcare delivery problem and the reasons for its study are provided. The discussion primarily centers around five key areas: the **Arthroplasty in Canada: Figures and Facts**, the journey **from pre-operative to post-operative** in arthroplasty, the **Literature review,** the adoption of **a multidisciplinary approach** for our research methodology**,** and the **thesis contributions** section that outlines key innovations of the proposed research.

1.1. Arthroplasty in Canada: Figures and Facts

Arthroplasty, a surgical procedure used to restore joint function [1], is a common and important treatment in Canada. Specifically, hip and knee replacements are among the most frequently performed inpatient surgeries in the country. Between 2020 and the end of 2021, there were over 44,000 arthroplasty operations conducted in Ontario alone [2]. Hip and knee replacements are particularly popular because they address joint issues, which are prevalent among Canadians [3]. In fact, in 2020-2021, there were 55,300 hip replacements and 55,285 knee replacements performed nationwide. Osteoarthritis was the leading cause for hip replacements, accounting for 69.4% of cases, while acute hip fractures made up 26.4% of the total. Knee replacements were predominantly carried out for osteoarthritis, comprising 99.3% of cases. The surgeries came at a cost of $1.3 billion, with a slight decrease from the previous year due to a 20.2% reduction in procedures (because of the Covid-19 pandemic). However, the average estimated cost for hip and knee replacement hospitalizations increased by 15.9%, amounting to $12,223[4] [5].

Hip and knee replacements not only rank high in terms of volume but also significantly improve patients' quality of life. A study conducted in the UK utilized the EuroQuol five-dimension (EQ-5D) questionnaire to assess the quality of life of patients awaiting these procedures. The results were staggering, revealing that 35.0% of individuals awaiting total hip arthroplasty and 22.3% awaiting knee arthroplasty scored below zero on the EQ-5D scale, indicating a state deemed "worse than death" [6]. Fortunately, post-operative assessments using the EQ-5D questionnaire demonstrated that 88.6% of hip replacement patients and 80.8% of knee replacement patients reported an increase in general health [7]. These scores were the highest among elective inpatient surgeries, surpassing procedures such as varicose vein surgeries.

The Canadian healthcare system operates on the principle that medically necessary services should be universally covered. Although Canada is often recognized for having a single-payer system, the administration of healthcare is handled at the provincial level. Each province and territory are responsible for its health insurance plans, care planning and funding, and fee schedules with healthcare professionals [8]. The Canada Health Act, enacted in 1984, ensures universal coverage for eligible individuals and services, and provinces must adhere to its principles to receive federal funding. Hip and knee replacements fall within the realm of services covered under this act, highlighting the importance of their efficient delivery [9].

Healthcare spending in Canada amounted to approximately $331 billion in 2022, equivalent to $8,563 per Canadian and representing 12.2% of the country's GDP [10]. Most healthcare costs, around 70%, are funded through general tax revenues. Provinces and territories contribute 78% of the funding, while the remaining portion is provided by the federal government through the Canada Health Transfer (CHT) [11]. Private sector contributions, including private insurance and out-of-pocket payments, cover costs for services not covered by the public system, such as prescription drugs, eye care, and dentistry. Hospitals, physicians, and medications consistently account for the largest shares of healthcare spending.

The COVID-19 pandemic posed significant challenges to the Canadian healthcare system, resulting in the postponement of surgeries and procedures to prioritize urgent care for COVID-19 patients. Over the first 22 months of the pandemic, nearly 600,000 fewer surgeries were performed compared to 2019, with hip and knee replacements accounting for 48,000 deferred procedures. The reduction in surgical volumes created substantial backlogs, as procedures were postponed without subsequent recovery in throughput levels [4]. To address the backlog effectively, provinces must surpass pre-pandemic surgery numbers—an achievement that has been realized nationally in only three separate months since the onset of the pandemic [12].

Of particular concern are the persistently long waiting times for hip and knee replacements [12, 13]. The percentage of Canadians receiving knee replacement surgery within the recommended waiting time of 182 days has dropped to 50%, compared to the pre-pandemic rate of 70%. For hip replacements, the percentage within the recommended waiting time decreased to 57%, down from the pre-pandemic rate of 75%. These figures represent the lowest values among procedures with established benchmarks. Moreover, both the 50th percentile and 90th percentile waiting times have increased, with the latter reaching 411 days for hip replacements. These prolonged wait times persist despite increased surgical volumes and larger patient pools factored into the calculations. This underscores the urgent need for solutions aimed at increasing surgical capacity and reducing waiting times[12–15].

In conclusion, arthroplasty, particularly hip and knee replacements, holds immense importance within the Canadian healthcare system. Despite being highly effective in improving patients' quality of life, the COVID-19 pandemic has exacerbated challenges, leading to a significant backlog and increased waiting times. Addressing this backlog and reducing waiting times require comprehensive strategies to increase surgical volumes and improve efficiency in healthcare delivery. These measures are crucial for ensuring timely access to arthroplasty procedures and alleviating patients' suffering.

1.2. Arthroplasty at The Ottawa Hospital, from pre-operative to post-operative

For elective joint replacement surgery at The Ottawa Hospital (TOH), patients are referred either through their primary care physician or a central intake clinic for hip and knee arthritis where triage is done by an advanced physiotherapist. This is to ensure that the patient has completed a course of non-surgical management for their arthritic pain which would include such things as anti-inflammatories and physical therapy.

After consultation with the orthopedic surgeon and joint replacement surgery has been consented for, the patient is placed on the surgeon’s wait list that is organized based on time of consenting as well as priority level based on degree/severity of pain and disability varying from 6 weeks to up to 6-8 months. The surgeon is allocated several surgery days per month to perform surgeries within the public hospital which ultimately dictates the wait time for the patient’s surgery.

The surgeon’s office calls the patient and books the surgery approximately 6-8 weeks ahead of time so that the patient can prepare for surgery and ensure that they are medically fit to undergo the joint replacement. There is an anesthesia assessment 2 weeks ahead of the surgery as well as an education session with a physiotherapist to prepare their home and their recovery in the pre-assessment unit. Many patients have same day discharge. The patient arrives at the hospital 2 hours ahead of their surgery time in the surgical day care unit (SDCU). They receive pre-operative pain medications as well as antibiotics. They are then brought to the operating room (OR) suite to undergo surgery under anesthesia.

After surgery is completed, they are brought to the post anesthesia care unit (PACU) to be monitored for 1-2hrs i.e., pain under control, alert and oriented, anesthetic medication worn off, patients are then transferred back to SDCU to be mobilized with physiotherapy and ensure they are comfortable and can transfer independently back home. Nurses ensure with the family present that their pain is under control and that they can urinate on their own prior to being discharged home. Further instructions are provided in regard to self-monitoring for complications as well as when to follow-up with their surgeon.

In terms of team selection, nurses in the SDCU, OR suites and PACU are trained and specialized for their areas of work. In other words, a nurse working in SDCU is not necessarily trained/qualified to work in PACU. Nurses in the OR suite require specialized training that takes several months. The anesthesiologist is randomly assigned to the OR suite a day or 2 ahead of the surgery day and is not necessarily the same person that assessed the patient in the PACU.

1.3. Current challenges of 4-joint surgeries at TOH & thesis objective

The goal for The Ottawa Hospital (TOH) was to achieve the highest possible Surgical Success Rate (SSR), i.e., completing 4 elective joint replacement surgeries, specifically hip and knee replacements, within a span of 8 hours and without running into delays. The designated time frame for this endeavor was from 7:30 am to 3:30 pm. This target was set to avoid incurring overtime costs, which were estimated at $56.84 per minute, resulting in an annual expense of $570,000 within the Division of Orthopedic Surgery.

While SSR may not be a widely recognized term, it was chosen to be a key performance indicator (KPI) as it is indicative to the specific problem we are addressing, OR throughput. The cases we selected involve straightforward arthroplasty procedures that are easily detectable, yet they still require overtime hours for completion. We intentionally excluded cases where complications occurred or where surgeons were aware in advance that the procedures would take longer than usual. Additionally, we excluded cases where complications arose three months after arthroplasty to ensure that our selection focused on the fundamental issue of being unable to complete four simple and predictable arthroplasties on time, i.e. 8-hour window.

In fact, when considering all arthroplasty cases at TOH, the success rate is indeed less than 39%, indicating that TOH often incurs overtime costs on most days. The primary rationale for opting for a binary output, namely, the ability to accomplish a day's work within an 8-hour timeframe, is aimed at addressing the performance of healthcare workers. The essential expectation is that they should efficiently handle routine cases within the designated time, which, unfortunately, they often struggle to achieve.

It is crucial to note that despite the binary nature of SSR's output, we calculated overtime as a continuous function using regression, as detailed in Chapter 6. We explicitly demonstrated the financial impact of paying overtime for straightforward cases and illustrated that the resulting savings could potentially fund additional cases, not to mention the overall saved amount.

It is essential to emphasize the value of expediting surgeries when it does not compromise patient safety, as this reduces the time patients spend under anesthesia, subsequently minimizing complications and surgical time. In our study, we proposed modifications to the work environment's behavior that contributed to delays and suggested changes to the processes and procedures of arthroplasty surgeries to streamline overall operation times. Notably, we did not advocate for surgeons or other clinicians to work at a faster pace. Our focus was specifically on the non-surgical aspects of the procedure, particularly those intensified during the pre- operative and post-operative intervals, as depicted in Figure 1. These non-surgical aspects pertain to roles unrelated to medical treatment or testing but rather concentrate on processes surrounding surgical procedures. In contrast, clinical approaches involve well-established roles in the healthcare industry, such as medical practitioners providing direct patient care.

Certainly, outcomes for patients - beyond safety considerations- such as examining potential surgeries, duration of hospital stays, readmission rates, injuries, patient satisfaction, and other relevant factors can be explored. This investigation into the impact on SSR could be conducted in the future when such data becomes accessible. This kind of data would also contribute to the comprehensive cost-benefit analysis by aiding in the assessment and estimation of the overall costs associated with achieving a faster SSR from various perspectives.

Despite various attempts and strategies, such as dedicating a room solely for hip and knee surgeries, implementing benchmarks for each stage of surgery, and exploring parallel processing by separating anesthesia activities, TOH has been unable to surpass a current SSR of 39% (i.e., since 2012).

The **research aim** of my thesis work is to enhance the SSR of arthroplasty surgeries at TOH. This will be achieved by leveraging supervised machine learning and data-driven solutions to improve the overall process from preoperative to postoperative phases. Additionally, the **research objective** is to enhance the efficiency of the entire healthcare team involved in joint replacement surgeries (i.e. surgeons, nurses, and anesthesiologists). The focus is on achieving these improvements while maintaining the highest standards of patient safety and optimizing resource allocation. Below is a brief literature review on the investigation of machine learning applied in the context of hospital efficiency improvement.

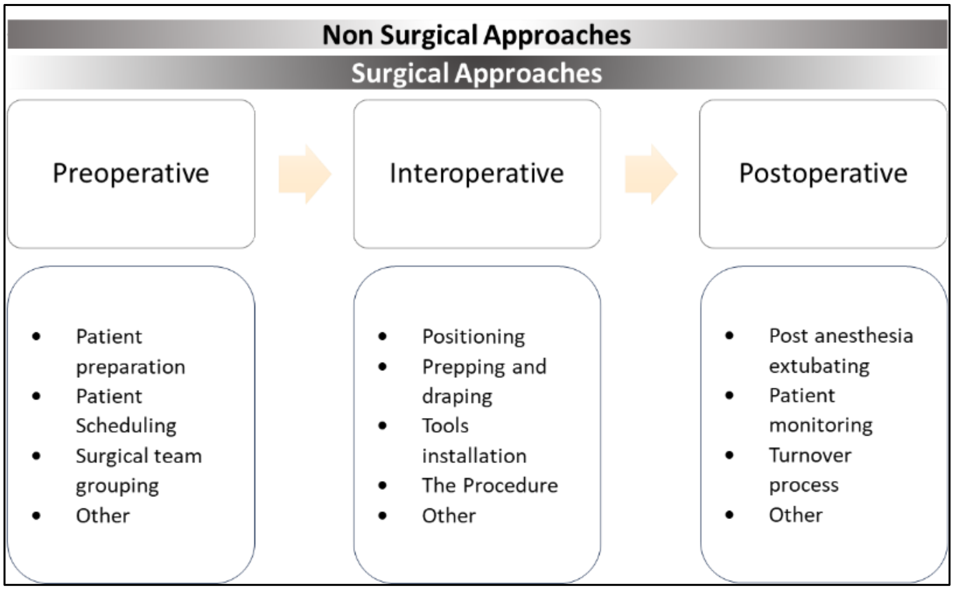


Figure 1: Non-Surgical approaches intensify efforts around the interoperative period

1.4. Literature review

Machine intelligence and operating room efficiency

Existing solutions that facilitate and optimize hospital efficiency are classified as either clinical or non-clinical approaches. Clinical approaches refer to the most recognized roles in industry. These encompass medical practitioners who offer direct patient care and typically dedicate numerous years to rigorous education and specialization in a particular medical field. Conversely, non-clinical approaches refer to roles that do not involve any form of medical treatment or testing [16]. They can be categorized into three facets: (i) introduction of new assets to improve hospitals' efficiency [17]; (ii) leverage of rhetorical data and descriptive analytics (data driven) to change, evaluate, or rearrange existing assets in a hospital, i.e., no machine learning involved [18]; and (iii) data driven solutions leveraging machine learning (ML) torniquets (prescriptive and predictive analytics). When specifically utilizing machine learning as shown in Figure 2, the efforts to improve efficiency can be directed towards scheduling patient surgeries, scheduling surgical team/resources, and optimizing surgical workflow time.

The work in the field of patient scheduling can be categorized based on the type of patient data utilized for modeling. Authors in [19] focused on geographical data, such as the distance of patients from the hospital and the potential impact of traffic on assessing the risk of no-shows or associated delays. They used this information to schedule patients and allocate operative days accordingly. In contrast, authors in [20] centered their work around sociodemographic data to extract patient characteristics. The scheduling data included factors like a patient's history of previous no-shows, the number of past appointments, the lead time for scheduling appointments, and whether the patient had insurance. Their machine learning scheduler predicted the likelihood of a patient not showing up and determined how many additional patients to schedule for a day to prevent the OR from being unused in case of cancellations. Another type of data found in the literature related to patient scheduling is time-related metrics that could influence patient decisions and delays. This category is exemplified in [21], where the authors considered metrics like patient waiting time, doctor idle time, and overtime. They proposed a system that predicts the likelihood of no-shows and then schedules appointments based on this risk assessment.

When it comes to surgical team allocation, there has been a substantial amount of work in utilizing machine learning for Human Resource (HR) allocation in various industrial sectors. However, the application of ML in healthcare has been relatively limited [22]. The predominant approach in HR allocation has involved the use of unsupervised ML techniques, particularly clustering. These efforts primarily focus on grouping individuals with similar characteristics, such as grouping patients with specific diseases [23]. Alternatively, some initiatives aim to create diverse groups with a wide range of capabilities, as seen in the case study groups for online education [24]. In clinics, a significant portion of HR allocation efforts has been dedicated to matching the right physician with the appropriate patients. This is done to address the demands of patients, optimize resource utilization, and ensure the delivery of high-quality healthcare in settings with limited resources [25, 26]. Additionally, there have been instances where ML has been employed to allocate teams and working hours during high-traffic periods, thereby preventing congestion, as demonstrated in [27].

Much of the workflow time-optimization solutions aim to predict a certain aspect of an operation; this can be predicting an event before surgery [28], after surgery [4, 29], or even the duration of the surgery itself [30, 31]. Concerning pre-surgical events, previous work usually attempted to develop a model that could successfully predict anesthesia preparation time. As for post-surgical events, authors addressed their hospital’s PACU crowding and predicted each patient’s PACU stay time based on their surgery. Additionally, other authors built a ML model that predicted the intensive care unit’s occupancy level in terms of number of beds. Finally, other solutions aim to establish benchmarks that help maximize efficiency or achieve set goals [31].

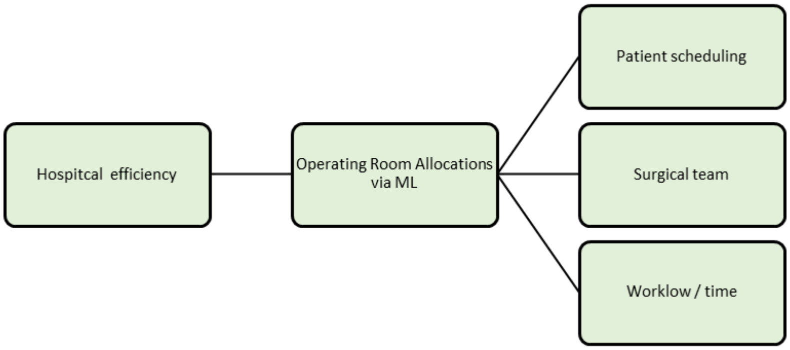


Figure 2: Overview of ML endeavors to optimise operating rooms efficiency.

Towards a Prescriptive Analytics System

There exists a multitude of definitions for prescriptive analytics (PA), and the prevailing commonality among these definitions is that the fundamental function of PA lies in its pivotal role of supporting decision-making, specifically by relying on machine-driven capabilities [32–36]. In my thesis, I define PA as the process of evaluating potential actions based on various outputs from predictive analytics to achieve the optimal solution for a specific problem.

Extensive literature has addressed the limitations of human cognitive abilities in solving complex problems. It has been observed that organizations that apply analytics to big data have achieved more accurate and improved outcomes in decision-making [37].

The Healthcare industry exemplifies a field that is increasingly driven by data due to the vast and expanding volume of information from various sources [38]. The most advanced healthcare analytics solutions excel when decisions require the evaluation of multiple alternatives [39]. Prescriptive analytics offers answers to questions such as "*what should we do to achieve certain goal*" based on insights generated by descriptive, diagnostic, and predictive analytics. Consequently, prescriptive analytics necessitates knowledge of potential actions to achieve better results. It employs what-if scenarios and causal relationships to optimize performance [37].
Throughout my thesis work, I created a generic blueprint for developing a Prescriptive Analytics System (PAS) [40] that aims to enhance the SSR rates for high-volume arthroplasty procedures. This blueprint will be introduced in the following sections.

1.5. The Multidisciplinary approach of my thesis

Among the 10 different key research themes offered by the School of Electrical Engineering and Computer Science (EECS), at University of Ottawa**,** I have decided to focus on the Systems and Machine Intelligence theme as shown in Figure 3. Machine Intelligence is an incredibly dynamic and multidisciplinary field that merges computer science, and cognitive science [41]. It holds immense potential for transformative applications across various domains of science, industry, and society. Professionals and researchers working in the field of artificial intelligence (AI) are increasingly recognizing the importance of adopting a multidisciplinary approach to their work [42].

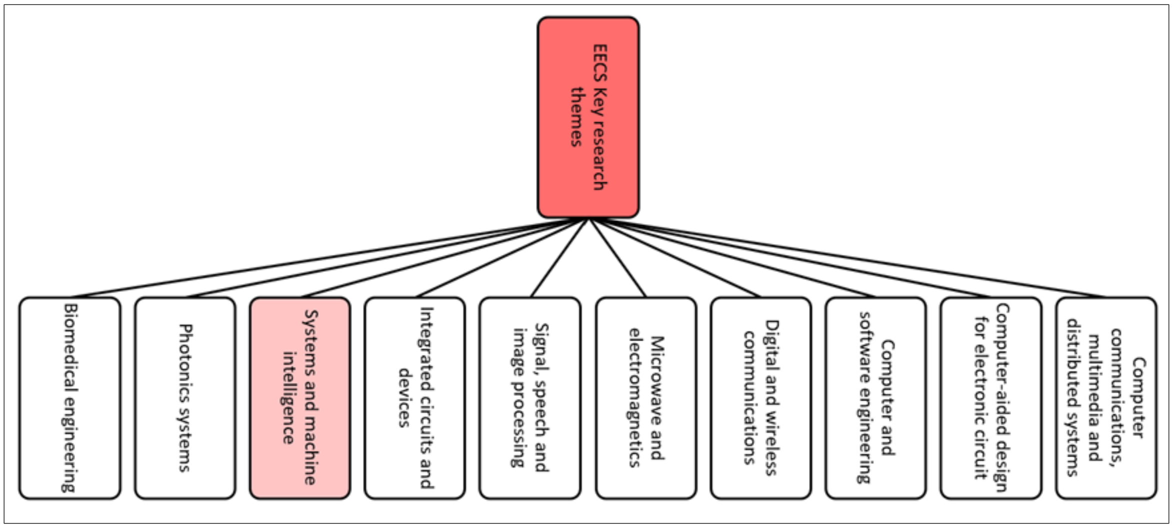


Figure 3: Key Research Theme Offered by the EECS at the University of Ottawa.

Within the engineering department at uOttawa, there exists a multitude of disciplines and research areas for exploration. In the context of my thesis work, I have directed my attention towards the axes highlighted in Figure 4, with a specific emphasis on the healthcare delivery sector.

The healthcare environment has undergone significant changes in recent times, characterized by higher patient acuity [43], cost-cutting measures [44], an upsurge in litigation [45], and elevated expectations from an educated generation of healthcare consumers [46].

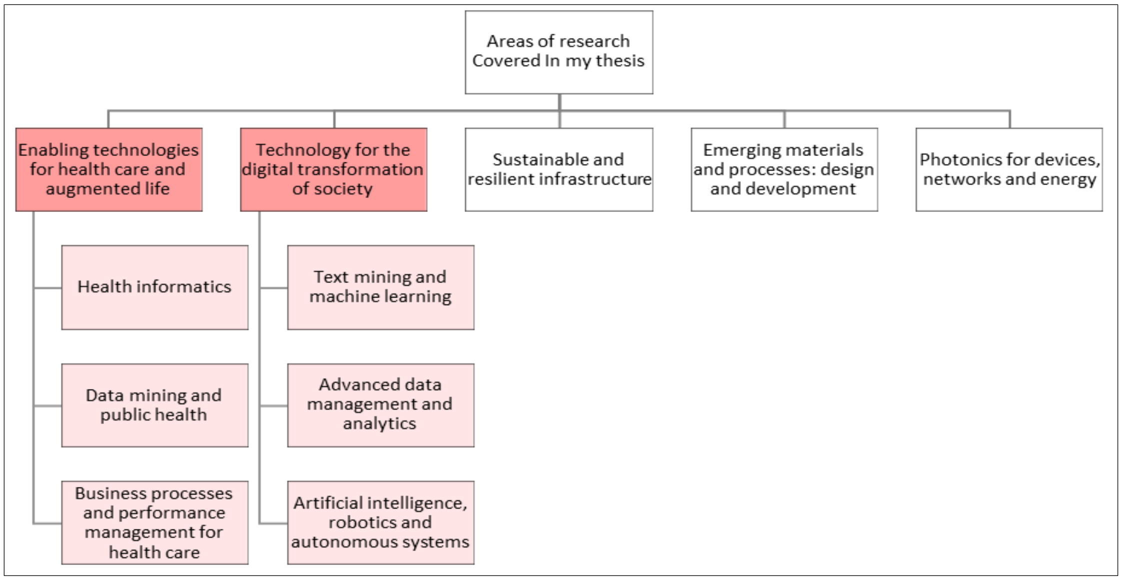


Figure 4: Research axes and multidisciplinary offered by the Engineering Department at uOttawa.

These factors have necessitated a continuous focus on measuring, assessing, and enhancing the quality of healthcare services. Quality improvement initiatives are not limited to patient clinical outcomes alone but also encompass customer service ratings and financial outcomes [47].

Achieving positive outcomes in quality improvement necessitates a collaborative approach, wherein building a cohesive and effective multidisciplinary team becomes crucial [48].

Recent advancements, particularly in the field of healthcare, offer promising prospects for the formation of multidisciplinary teams. Medical professionals now have access to AI-aided computing, which provides refined databases for healthcare prevention, treatment, and diagnosis. The emergence of AI has broadened the understanding of healthcare beyond its association solely with physicians, nurses, and other essential personnel. People are increasingly becoming aware of the remarkable implications and advantages of AI in the healthcare setting [48].

In the realm of Systems and Machine Intelligence, which falls under the School of Electrical Engineering and Computer Science at the University of Ottawa, my thesis revolves around the design, implementation, and validation of a Prescriptive Analytics System (PAS) based on machine learning. <u>The primary goal</u> is to <u>enhance the throughput of operating rooms</u>. Specifically, I proposed three frameworks, each integrating machine learning engines, with distinct types of input and varied output as well. Subsequently, **<u>one</u>** of these frameworks was chosen for conversion into a decision support system, complete with a dashboard designed to facilitate the utilization of both machine learning and framework outputs. I proceeded to implement this PAS in a clinical setting and conducted an evaluation of its performance.

The diverse nature of my PAS "a decision support system DSS" required contributions not only in EECS-related journals but also in clinical journals. Implementing the PAS in practical healthcare settings necessitated a collaborative effort to enhance knowledge in the healthcare domain, resulting in an integrated solution ready for real-life application.

In summary, my thesis work focuses on the healthcare delivery sector, addressing the need for quality improvement of arthroplasty procedures through the integration of AI within the Division of Orthopedics Surgery, at TOH. The blueprint for my proposed healthcare delivery AI-transformation is summarized in Figure 5.

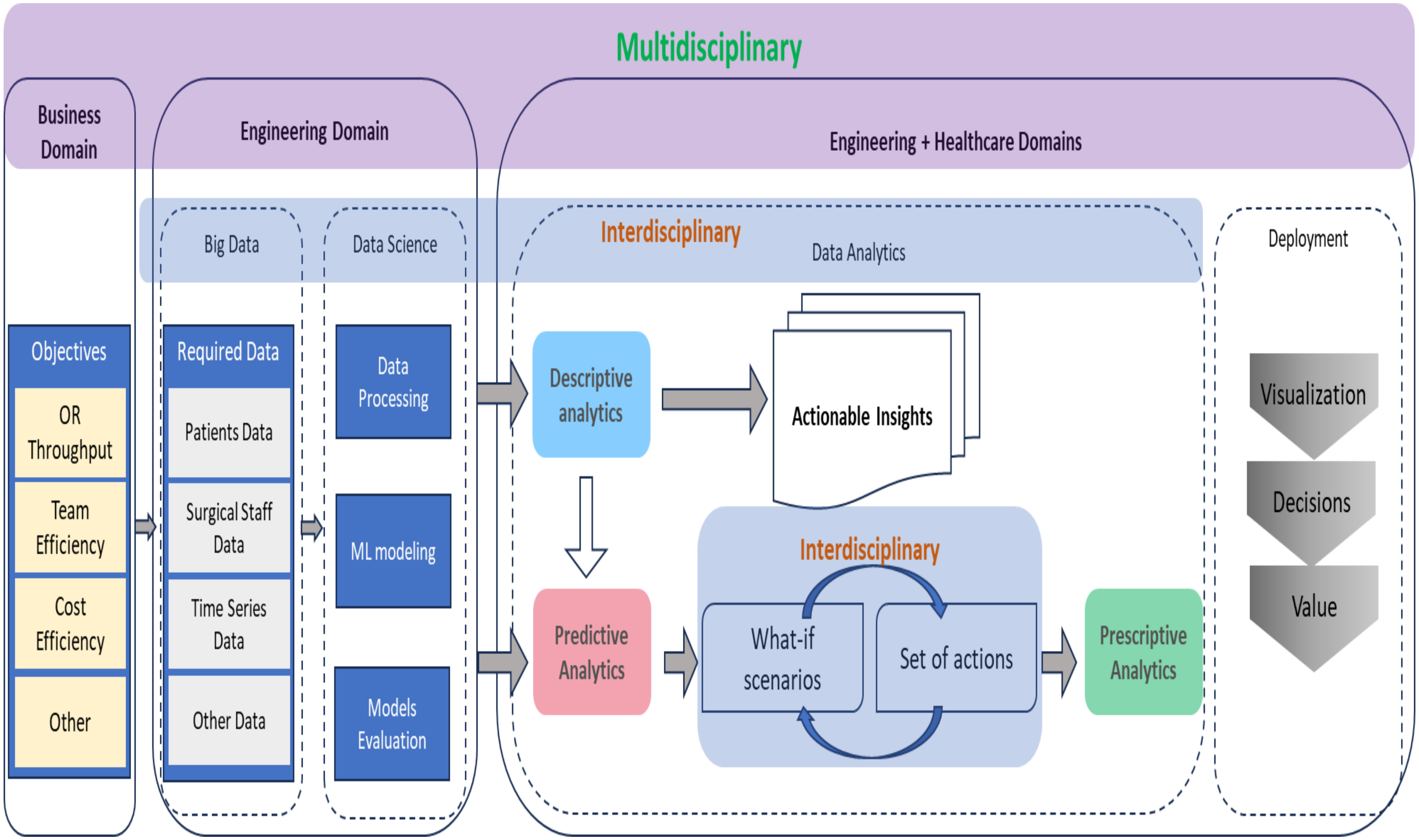


Figure 5: The adoption of multidisciplinary approach and prescriptive analytics for AI-Transformation for arthroplasty surgeries at TOH.

1.6. Thesis Contributions

Figure 5 depicts the schematic of my proposed AI-driven transformation for Arthroplasty surgeries, while also highlighting the milestones and research contributions of my thesis. As the thesis format is a “thesis by articles” one, I now highlight the relevant contributions to their respective publications.

The initial phase of my thesis focused on descriptive analytics (DA). I analyzed structured data to understand simple relationships, correlations, and trends among arthroplasty surgery metrics. This analysis led to the publication of my first journal titled "***Factors Influencing Delays and Overtime During Surgery: A Descriptive Analytics for High Volume Arthroplasty Procedures***." The data was categorized into three groups: time-related metrics, patient-related metrics, and staff-related metrics. DA helped gain insights into the surgical process and identified standalone metrics that impact surgical delays.

I then proposed three frameworks for each category of data: patient scheduling framework, surgical team scheduling framework, and workflow/time monitoring framework. These frameworks were presented in my second journal, "***Frameworks for AI-based Surgical Transformation (FAST)***" where I compared various supervised machine learning models within each framework, evaluating their complexity and implementation aspects. I have the option to use these frameworks either in combination or independently. Nevertheless, I have opted for the one that offers the greatest expected value to fulfill the objectives of enhancing both OR throughput and team efficiency during the transformation of our Prescriptive Analytics System.

As discussed in the literature review, to construct a Prescriptive Analytics System, it is essential to gather the necessary potential actions for achieving improved outcomes. This is accomplished by utilizing what-if scenarios and establishing causal relationships that guarantee the desired performance. Those actions and rules that can be implemented to achieve better SSR are outlined in my journal "***Use of multidisciplinary positive deviance seminars to improve efficiency in a high-volume arthroplasty practice: a pilot study***".

The final component needed to fully depict the Prescriptive Analytics System landscape, comprising descriptive insights, the prediction framework, and the essential actions, entails integrating them into a functional prototype within a workflow system. This integration should yield practical results, encompassing both optimistic and pessimistic scenarios, as detailed in my journal titled “***Artificial-Intelligence driven prescriptive model to optimize team efficiency in a high-volume primary arthroplasty practice***”. The article revolved around the process of defining benchmarks standards for individual surgical stages, prioritizing important metrics while omitting less significant ones. It utilized a dataset of around

5000 patient surgeries to make predictions, followed by the automatic generation of a AI-recommended set of actions aimed at achieving predefined Surgical Success Rates.

My article, "***Leveraging Machine Learning and Prescriptive Analytics to Improve Operating Room Throughput***" delved into the practical implementation of the PAS. It elucidated the system's applications in real-life scenarios and explained its functionality under different PAS output scenarios.

Finally, the "***First Deployment of Artificial Intelligence Recommendations in Orthopedic Surgery***" is my final journal. The article disseminates the first results and knowledge of the PAS on 200+ arthroplasty surgeries at TOH.

I have consistently referenced my journals throughout this thesis using abbreviations J1 to J6 in the same order as listed above. Additionally, Table 1 employs these same abbreviated names for clarity. Similarly, my abstracts have been designated as A1 and A2 for simplicity.

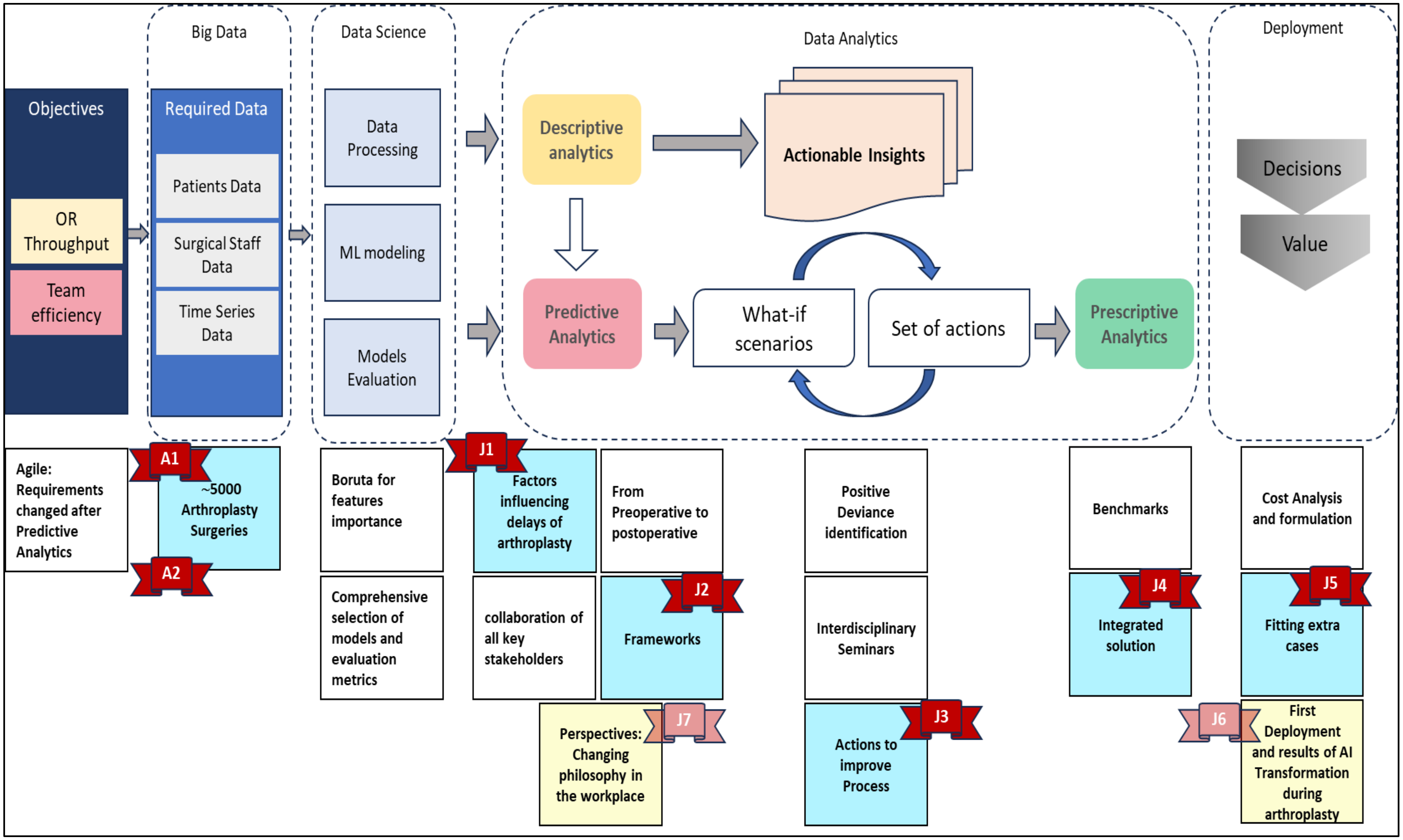
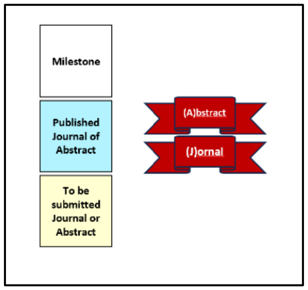


Figure 5: Schematic of the AI-driven transformation of the Arthroplasty Operating Room, while highlighting the milestones that mark



Chapter 2: Factors Influencing Delays and Overtime During Surgery: A Descriptive Analytics for High Volume Arthroplasty Procedures

2.1. Summary

This journal aimed to gain insights into the business of surgery by conducting descriptive analytics on the available structured data. The primary objective was to understand the underlying factors contributing to delays and overtime pays in arthroplasty surgeries, and how these factors impact the SSR at the Ottawa Hospital. The approach involved analyzing the data through visualizations and simple central tendency metrics to uncover meaningful patterns and relationships. By delving into the data, we aimed to extract valuable information and gain a comprehensive understanding of the surgical process and its impact on performance indicators.

2.2. Methodology

The data mining process, following the guidelines of the Cross-Industry Standard Process for Data Mining (CRISP-DM), commences with a focus on business understanding. In our efforts to comprehend arthroplasty surgeries, our initial step involved conducting descriptive analytics on the available data. We sought to establish the relationship between each metric and SSR individually, marking the first level of analysis. Subsequently, we explored all possible combinations (excluding time stamps, only one level of combination) of two and three different metrics against SSR in what we termed level-2 and level-3 analyses, detailed in my descriptive analytics article.

The significance of metrics, their meanings, and the potential influence of each metric on SSR were initially discussed with field experts, including clinicians, surgeons, and statisticians. The outcomes of our descriptive analytics primarily comprise observations and insights that, if implemented, would incur minimal costs (financial, labor, and impact) with a high perceived gain according to expert opinions. However, we refrained from making any changes at this stage until we compared these findings with the outputs of the machine learning and decision support systems.

It's worth noting that descriptive analytics has its limitations, leading us to rely on the machine learning-based system in constructing our Decision Support System (DSS).

While the primary aim of this analysis was to grasp the surgical process and factors influencing arthroplasty delays, statisticians were consulted, and they indicated that statistical tests were deemed unnecessary or unfeasible in the majority of our descriptive analytics outcomes we have. Nevertheless, for the purpose of quantifying important metrics, I included some additional, to the ones in the published article, statistical analyses and tables in Appendix A where applicable. Additional facets were addressed in the appendix concerning various approaches to compute SSR for both surgeons and patients.

2.3. Fundamental research contribution

First published analysis of descriptive analytics to identify factors contributing to low SSR at TOH.

2.4. Author Contribution

Author	Contribution to the study
Farid Al Zoubi	Data curation, Investigation, Methodology, Writing - original draft, Validation, Writing – review and editing
Paul Beaulé	Conceptualization, Methodology, Writing – review and editing
Pascal Fallavollita	Conceptualization, Methodology, Funding acquisition, Project Administration, Resources, Supervision, Writing – review and editing

2.5. The Article

The following article was submitted to the Frontiers in Surgery Journal and is currently under review.

frontiers | Frontiers in Surgery

TYPE Original Research
PUBLISHED 04 January 2024
DOI 10.3389/fsurg.2023.1242287

Factors influencing delays and overtime during surgery: a descriptive analytics for high volume arthroplasty procedures

Farid Al Zoubi[1]*, Paul E. Beaulé[2] and Pascal Fallavollita[3]

[1]School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, ON, Canada,
[2]Division of Orthopedic Surgery, Ottawa Hospital Research Institute, Ottawa, ON, Canada,
[3]Interdisciplinary School of Health Sciences, University of Ottawa, Ottawa, ON, Canada

Check for updates

OPEN ACCESS

EDITED BY
Vassilios S. Nikolaou,
National and Kapodistrian University of Athens, Greece

REVIEWED BY
Roy Eagleson,
Western University, Canada
Stefano Marco Paolo Rossi,
Fondazione Poliambulanza Istituto Ospedaliero, Italy

*CORRESPONDENCE
Farid Al Zoubi
✉ falzo100@uottawa.ca

RECEIVED 18 June 2023
ACCEPTED 06 December 2023
PUBLISHED 04 January 2024

CITATION
Al Zoubi F, Beaulé PE and Fallavollita P (2024) Factors influencing delays and overtime during surgery: a descriptive analytics for high volume arthroplasty procedures.
Front. Surg. 10:1242287.
doi: 10.3389/fsurg.2023.1242287

The aim of this article is to analyze factors influencing delays and overtime during surgery. We utilized descriptive analytics and divided the factors into three levels. In level one, we analyzed each surgical metrics individually and how it may influence the Surgical Success Rate (SSR) of each operating day. In level two, we compared up to three metrics at once, and in level three, we analyzed four metrics to identify more complex patterns in data including correlations. Within each level, factors were categorized as patient, surgical team, and time specific. Retrospective data on 788 high volume arthroplasty procedures was compiled and analyzed from the 4-joint arthroplasty operating room at our institution. Results demonstrated that surgical team performance had the highest impact on SSR whereas patient metrics had the least influence on SSR. Additionally, beginning the surgical day on time has a prominent effect on the SSR. Finally, the experience of the surgeon had almost no impact on the SSR. In conclusion, we gathered a list of insights that can help influence the re-allocation of resources in daily clinical practice to offset inefficiencies in arthroplasty surgeries.

KEYWORDS

descriptive analytics, health care efficiency, high volume surgery, operating rooms throughput, overtime hours

1 Introduction

The term arthroplasty is the amalgamation of Arthro (Greek), meaning joint, and Plasty, which means to mold, graft, or reform. Hence, arthroplasty is the science of molding or reforming a joint, usually to reclaim its full function or relieve joint pain (1). Human joints become painful and stiff (with age) from regular wear and tear alone, but certain degenerative diseases can exacerbate the condition. Arthroplasty surgeries are the ultimate corrective measure to rectify these conditions. The procedure may include retaining the healthy parts of the joints and augmenting them with implants (i.e., resurfacing, partial replacement procedures) or completely replacing the joint (both ends) with implants.

To address issues with hospital efficiency, various initiatives to increase throughput, such as high-efficiency operating rooms (ORs) and parallel processing with anesthesia block rooms, have been suggested (2). At our hospital, we instituted increased through put rooms going from two-three to dedicated four primary joint rooms with dedicated arthroplasty surgeons in. The 4-joint OR was designed specifically to handle these procedures, and everything, from its layout to equipment, has been arranged with arthroplasty surgeries in mind. The design allows surgeons and staff to save time on procedures and complete more surgeries in a day than a general OR handling different types of surgical patients would allow for.

In the area of arthroplasty, there exists few works that identify factors which influence surgical outcomes. Authors in (2) determined that patient length of stay is multifactorial and can be reduced by regular review of the care pathway to effect incremental changes that have been identified as having an impact on reducing stay. In (3), authors identified specific factors that ensure positive patient outcomes following knee surgery, both non-surgical (i.e., gender, age, body mass index, etc.) and surgical factors (i.e., anesthesia, postoperative complications, and rehabilitation). Lastly, authors in (4) concluded that patients' perception of pain control was significantly positively correlated with the perception of their orthopedist, nurse, and overall hospital satisfaction.

Different to the state-of-the-art, the objectives of our study are to identify the factors which influence Surgical Success Rate, or SSR, which is the ratio of successful surgery days over total surgery days. To our knowledge, this is the first attempt at identifying factors which contribute to surgical delays/overtime in the application area of orthopedics. The definition of successful surgery days (in this context) is a day in which all four joint surgeries scheduled are completed within the dedicated eight hours (between 7:30 am and 3:30 pm). An unsuccessful day has two negative consequences:

1. Overtime, which costs our institution $570,000 a year (5). The dollar amount was calculated by multiplying the number of additional minutes an OR was engaged for (during unsuccessful days = 10,179 minutes) by $56, which is the per-minute collective cost of an OR and the staff using it.
2. Postponing the fourth patient of the day to a future date. This results in backlog, low patient satisfaction rates, and underutilization of hospital resources with the third case often ending at 14:30–15:00, let alone the unused time for days that team manages to complete their fourth case before 15:30. This unused time is calculated to be an average of 36 min per successful day in our institution.

The following sections of this work are divided into three levels based on the number of surgical variables being analyzed simultaneously, for which level one is the simplest format of analysis while level three is the most complex analysis.

2 Descriptive analytics

Descriptive Analytics (DA) is the science of analyzing available data to determine patterns and trends. It focuses on "what happened?", not how, why, or if it might happen again. It's relatively easy to understand, which makes it useful and accessible to a wider audience. It can offer a wealth of useful insights and help with decision-making (6).

The primary advantage of descriptive analytics is that it allows you to view how certain variables, relationships, and trends change over time. This, along with its simplicity, makes it quite useful to evaluate and communicate performance. Anything that can be quantified via metrics, changes over time, and has discernable patterns, is in the realm of descriptive analytics. Another advantage it offers is easy-to-comprehend visualization of complex numerical data, which makes it more palatable for people from different departments and disciplines. More eyes on data can help with more insights and unique perspectives, which aid in decision-making (7).

Descriptive analytics has multiple use cases in healthcare (8). It's used for trend analysis, such as identifying which age group and BMI category has the most joint replacement surgeries (9). It also assists in planning, such as stocking up on flu vaccine before certain months of the year, based on past trends (10). Descriptive Analytics can also lead experts to the right causes behind certain trends just by helping them realize what to look for.

Multiple types of descriptive analysis techniques and approaches are associated with both individuals and clinical units (7), including case reports (11), cross-sectional studies (12), and surveillance (13). For our study, we are following the passive surveillance approach for descriptive analytics, i.e., systematic collection of data pre-divided into formal categories and spread out over an adequate period of time (i.e., enough for cyclical patterns to emerge).

The retrospective data we have compiled and analyzed for this study comes from the 4-joint operating room for arthroplasty procedures at our institution. The critical characteristics of the data are as follows:

- Time Period: 2012–2020
- Nature of Procedures: Non-complex cases, Unilateral hip and/or knee replacement surgeries only.
- Nature of Data: Numerical (mostly time stamps and time durations) and Categorical
- Data Collection Source: Surgical Information Management System (SIMS) for the majority of the metrics, while some other were collected manually from daily notes.
- Data Dimensions: 26 Columns and 788 rows. See **Supplementary Table S1** for all metrics.
- Treatment of Data: Data was cleaned for missing information and incorrect values, both of which represented less than 1% of the observed dataset. Rare cases, regardless of their dissent with median values, were kept in the dataset.
- Surgical Success Rate (SSR): was calculated at 39%.
- Demographics and Other Quantifiable:
    - Number of days and surgeries: 197 days (788 surgeries)
    - Number of surgeons: 6
    - Number of Nurses: 73
    - Number of Anesthesiologists: 81
    - Gender-wide Distribution of Sample Patients: 385 M, 403 F
    - The average age of patients: 63.2 ± 11.9
    - Average BMI: 30 ± 5.7

Time intervals used are a modified version of those defined by the Association of Anesthesia Clinical Directors (AACD): anesthesia preparation time (APT); patient in-room to anesthesia ready, surgical preparation time (SPT); anesthesia ready to procedure start, procedure; procedure start to procedure finish, anesthesia finish time (AFT); procedure finish to patient out of room, and turnover time; and first patient exits to subsequent patient in room. APT immediately follows turnover. Figure Below illustrates the surgical intervals along with their spans.

3 Level-1: SSR vs. individual metrics

At Level-1, we are analyzing each metric individually and how it may influence the SSR. Analyzing SSR from the perspective of each metric can help us identify outliers, irrelevant factors, and trends that may otherwise get buried under the data. Another benefit of focusing on individual metrics is the ability to weigh each metric for its influence on SSR or, at least, identify metrics with the most significant and least significant impact on SSR, **Supplementary Table S2**. Less resource-intensive and high-impact metrics can help us develop intervention strategies that may directly reflect in a high SSR. Conversely, more resource-intensive, high-impact metrics can become the elements of a more comprehensive, long-term strategy to improve the success rate.

We have divided the metrics of Level 1 of analysis into three different categories:

1. Staff and facility-specific metrics: calculating the impact of controlled and managed resources (i.e., primarily human resources) on SSR.
2. Patient metrics: calculating the impact of these metrics on the SSR offers great insights regarding patient management and scheduling, especially with the data/insights we have access to from the first category.
3. Time-related metrics: identifying which aspects of the surgeries have the most significant impact on the success rate and their timely completions. This helps with the development of multi-faceted optimization strategies covering both individuals and processes to improve SSR.

3.1 SSR vs. staff and facility

3.1.1 SSR vs. campus

Our institution has two campuses: Civic and General. The bulk of the surgeries happen at the General; with a 7:1 ratio of surgeries. There is only a 4% difference in the SSR, as 36% of the Civic Campus's surgeries are successfully compared to 40% at the General Campus, average SSR is 39%. The difference is not nearly as significant as the difference in the number of surgeries, and no other metric supports the assumption that a higher number of surgeries resulted in a higher SSR (see **Supplementary Figure S2**).

3.1.2 SSR vs. surgeon

The SSR varies greatly from one surgeon to another. The lowest extreme is 13.3% SSR (if we neglect the seventh surgeon with a 0% SSR), and the highest extreme is 62.3% (see **Supplementary Figure S3**). One curious observation from this comparison is that there is no discernable connection between the number of surgeries and SSR, i.e., SSR is not tied to the experience gained from performing more surgeries. Surgeons S3 and S4 have a minimal difference in the number of surgeries they conducted, but the SSR difference is significant 41.1% vs. 62.3%. This is further endorsed if we compare the three closest SSRs for surgeons S4, S5, and S6. S6 completed roughly 4.6 times, and S4 completed 6 times more surgeries than S5, but the SSR difference is minimal. This is also not a true reflection of a surgeon's capabilities, at least not without taking other factors like team and surgery type into account. However, it may help identify the best performer and worst performers if their difference from the mean is significant enough.

3.1.3 SSR vs. anesthesiologist

Like the result above, the SSR is not correlated to the number of surgeries an anesthesiologist has been a part of. In fact, the opposite is more plausible, i.e., the higher the number of surgeries, the lower their SSRs might be (**Supplementary Figure S4**). This is supported by the fact that there is just one anesthesiologist with a 100% SSR that completed more than ten surgeries and **Supplementary Figure S5**: Circulating nurse's experience does not influence SSR.at least fifteen anesthesiologists that completed less than ten surgeries. The anesthesiologists with high SSRs can be considered a controlled factor for future surgeries to influence the probability of a surgery succeeding on time. However, to determine the potency of this controlled factor, it's imperative to take the influence of an anesthesiologist on a surgery completed on time into account.

3.1.4 SSR vs. circulating nurse RN1 and RN2

The number of surgeries a registered nurse (RN) attends to does not influence the SSR, and the fewer surgeries an RN has attended to, the higher their chances of achieving a respectable SSR. However, it's difficult to identify discernable trends because of the statistical weaknesses of this dataset or, more accurately, its distribution (**Supplementary Figure S5**). The top 3 RN2s and top 4 RN1s have completed more surgeries than the rest combined. The uneven distribution of sample data makes it impossible to identify the connection between a RN and SSR.

3.2 SSR vs. patient metrics

A patient's physical condition, the type of surgery they need, their age, and gender can have a significant impact on the successful completion of surgery on time.

3.2.1 SSR vs. sex

The SSR for Male patients is slightly higher than for male patients (**Supplementary Figure S6**). It's consistent with the finding of another study that investigated the operative times of surgeries for male vs. female arthroplasty patients. A study has demonstrated that men are at higher risk of developing prosthetic joint infections following joint arthroplasty, thus surgeons have to take extra precautions during surgeries (14). However, we believe that the real practical reason is that men are more muscular which makes surgery more difficult. This insight can be used for patient scheduling to improve the overall SSR. Scheduling two males and two females per day, or scheduling three or four females in one day when there are surgeries with anticipated complications, can be strategically helpful to make up for time delays and enhance the chances of completing four surgeries in the allotted time.

3.2.2 SSR vs. age

The bulk of the age-wise SSR trend hovers between 32% and 46%, with one outlier being the age group between 17 and 26. The youngest group also has the lowest number of surgeries, and it's consistent with typical age-oriented surgical recovery and success trends. But apart from that, there is no discernable trend. There is an 8% difference in the SSR for people between the ages of 57 and 66 and patients between 27 and 36 years of age, with older patients having a higher SSR (**Supplementary Figure S7**).

3.2.3 SSR vs. BMI

The BMI correlation with SSR offers pattern abnormalities similar to age (**Supplementary Figure S8**). It's highest for patients in the Class 3 obesity BMI. This is inconsistent with the observation for both elective surgeries like Total Knee Replacement (TKA) and Total Hip Arthroplasty (THA).

3.2.4 SSR vs. ASA

The primary concern with identifying patterns when comparing SSR with the American Society of Anesthesiologists (ASA) physical status classification system classes is the data distribution. The sample sizes of Class I and IV are lower compared to Class II and III. If we average out Class II and III (about 367 cases), Class I is 10.3%, and Class IV is 3.4% of that sample size. Between two reasonably comparable classes (II and III), the pattern is as expected—higher for a safer ASA class and lower for a riskier class (see **Supplementary Figure S9**).

3.3 SSR vs. time stamps

3.3.1 SSR vs. months

An interesting pattern was observed when we analyzed SSRs for different months of the year. Apart from two exceptions (July and August), the remaining ten months can be divided into sets of two. Five of them are above 50%, and five are between 30% and 40%. May, the month with the highest SSR, is a true outlier, and August, the month with the lowest SSR, is the culmination of a four-month-long downward trend. The variation in the number of surgeries for each month is also a pattern worth considering, as it may be tied to factors like staff availability and fatigue (15). However, it doesn't impact the surgical success rate as both the highest and lowest SSR months had only a difference of about ten surgeries, which is less than 15% of total surgeries for either month (**Supplementary Figure S10**).

3.3.2 SSR vs. days

The SSR for days shows that the best days for surgery are one day after the weekend ends and one day before the weekend begins, i.e., Tuesday and Thursday. Also, for days, the pattern of more surgeries resulting in a higher SSR holds apart from one outlier (Monday). This could be construed that it's tough to get to work on Monday, and on Fridays' the majority of people are looking forward to the weekend which may influence their focus on surgery (**Supplementary Figure S11**). One study shows that employees are less supportive on those days, i.e., Monday and Friday (16).

3.3.3 SSR vs. time in room and anesthesia ready time

The correlation between Time in Room and Anesthesia Ready Time is evident from the SSR pattern for both variables, and it's tied to the starting time of the surgery (**Supplementary Figure S12**). The SSR is higher for surgeries in which the patient was in the room and anesthetized closer to the scheduled time/allotted time slot. The farther away they were from that time window, the lower their SSR became. For example, if by 8:20 AM (Cut-off Time) the first patient (P1) was not in the OR already, there is no way the fourth surgery can be completed without having to pay overtime. Another example is that if the third patient (P3) did not have his anesthesia ready by 1:00 PM, there will be a very slim chance (less than 20%) the fourth surgery would be completed on time, i.e., before 3:30 PM. As an observation, the nurses arrive to work at 7:30am and there is no real accountability for that first 25–29 min in terms of productivity as long as the patient is in the room before 8:00 am and as one would expect this often spill into after 8:00 am.

3.3.4 SSR vs. case start and case finish time

A similar pattern was observed when we compared SSRs against Case Start and Finish times. The procedures that started and ended in the allotted time slots had a much higher SSR rate. The four waves in **Supplementary Figure S13** represent four-time slots for four arthroplasty surgeries in a given day, along with the cut-off time for each wave where the patient should be no later to consider it as a successful day.

3.3.5 SSR vs. time out of room and anesthesia stop time

The pattern is the same for SSRs when compared against Time Out of Room and Anesthesia Stop Time—four waves endorsing the observation that surgeries that start on time and end on time resulted in higher SSR days (see **Supplementary Figure S14**).

3.3.6 SSR vs. anesthesia start time and turnover

The anesthesia start time doesn't conform to the same pattern, at least not with the same degree of correlation, as other time metrics when compared to SSR. A downward pattern is observed between SSR and turnover rates (i.e., the time to prepare the room for the next surgery), but it includes a hard spike and a relatively hard slump. On the other hand, turnover is measured in minutes, thus **Supplementary Figure S17** represents duration rather than time stamps. Turnover time should take between 12.9

and 17.9 min to achieve the highest SSR, considering the 42.9–47.9 window as an outlier since there are not enough samples to support this high SSR.

4 Level-2: comparing timestamps vs. patient metrics

As we analyze a level deeper, we are comparing two to three variables at once. It's mostly a three-dimensional analysis compared to Level-1, where we compared one to two variables/metrics side by side to determine a pattern. It's also different from the Level-1 time-metric comparisons because it compares averages to patient metrics instead of timestamps.

4.1 Time vs. age

The average procedure time and the average case total time follow an almost parallel pattern since procedure time makes up the bulk of the case total time. Interestingly, the turnover rate follows a similar pattern. The Anesthesia Preparation time (APT) average is the most obvious outlier, as the average consistently goes up until the second last age group and then drops off. If we observe the averages excluding the two extremes, it's clear that Anesthesia preparation and in-room time go up with age. For patients above 45 years old, the chances of spending more time in the OR become higher as they reach 76 years old and steadily lower after 76 years old (**Supplementary Figure S16**).

4.2 Time vs. BMI

The averages for APT in Room, Surgery Finish time (SFT), and turnover have gone up with the BMI. In contrast, averages for Surgery Preparation Time (SPT) and procedure went down as BMI increased. The Case Total average time is most significantly influenced by the procedure average time and APT average, which rises sharply with BMI but drops off for the riskiest BMI class (**Supplementary Figure S17**). It also shows the APT's influence on Case Total, which followed the APT's trajectory instead of the procedure averages, between the BMI of 27.1 and 47.1.

4.3 Time vs. sex

On average, surgeries for male patients takes 5 min longer than surgeries for female patients. The SPT and SFT averages for males are also slightly higher (one minute on average), which pushes the total time difference (Case Total) to six minutes (**Supplementary Figure S18**). This is one rationale behind the higher SSR for female patients as shown in **Section 3.2.2**.

4.4 Time vs. ASA

Anesthesia-related averages are following the naturally expected pattern, i.e., moving up for higher/riskier ASA classifications, though there is virtually no difference between Class I and II. The pattern for the average Case Total and Procedure is not influenced by the natural ASA pattern. In fact, it's going in the opposite direction (**Supplementary Figure S19**).

4.5 Time vs. type of surgery

The Case Total average time is inversely related to SSR. The HRA, with the highest SSR, takes the most time, and UKA takes the least amount of time. However, the variance in time is not nearly as significant as it is for SSR when it comes to different types of surgeries. There are significant similarities between the two knee surgeries and two hip surgeries, respectively. The only outlier is the average APT which is significantly higher in TKA (**Supplementary Figure S20**). The APT average for hip procedures is significantly lower compared to knee operations but has minimal to no impact on the Case Total average. Procedure time and APT in Room may have the most significant influence on the Case Total average.

5 Level-3: staff vs. patient vs. time-metrics analysis

At the highest level, we are analyzing four variables at once to identify more complex patterns in data and correlations that are invisible or not credible enough at Level-1 or Level-2. We are analyzing two-time metrics and one staff/Patient metric with SSR. Different four-variable combinations can help us identify a wealth of insights and trends via a comprehensive descriptive analysis.

5.1 Surgeon-SSR vs. procedure and SPT averages

We are comparing surgeons' (with their respective SSR) numbers for their procedure time average (highly relevant to surgeons) and SPT (less relevant to surgeons). In this scenario, one extreme would be the surgeon with a high SST and low procedure average, and the other would be a high procedure time and low SST. Surgeon PB is an example of the positive extreme, but they have also benefited from low average SPT. Surgeon GD is an example of the other extreme who, despite having low SPT, had high procedure times and low SSR (**Supplementary Figure S21**). Surgeons (with their respective SSRs) were plotted for the following (X and Y axis) variables:

- SFT and Average of AFT (Clustering—SFT 3–6 and AFT 9–16)
- AFT and APT (Clustering—AFT 20–40 and APT 9–16)
- APTinRoom and SPT (Clustering—APTinRoom 10–16 and SPT 13–16)

However, no discernable pattern was observed, apart from clustering in certain intersections of the above-stated variables. Recall that it was the same for the circular nurses when they were plotted against turnover, APTinRoom, and SFT.

5.2 Surgeon-SSR vs. BMI and age

The most successful surgeon (based on procedure time and SSR) has operated on patients with the lowest average BMI and age of all surgeons. However, this is not the case for the other extreme. Age seems to have a far more significant impact on the average time a surgeon takes to complete a procedure than BMI. However, the limitations of this analysis should be considered with reference to the sample size (more surgeons to compare). A greater sample with multiple data points concentrated within a specific age range (like below 60 or above 75) can cast a shadow on the strength of this correlation (**Supplementary Figures S22**, **S23**).

6 Discussion

Descriptive analytics (DA) helps us decipher raw data from the past and identify patterns and trends to generate useful insights that may be applied to future decision-making. By identifying relationships between different metrics (variables), it helps us differentiate the most crucial metrics from relatively non-important ones that may not have a significant enough impact on trends and SSR. Comprehensive DA and the identification of the most important metrics can become the foundation for more advanced Diagnostic Analytics, which focuses on the reasons and rationales behind certain trends, i.e., the "why" behind what happened (14).

Understanding how different metrics/variables interact with each other and how they impact SSR and metrics that directly influence SSR (Time Metrics: Complete Case Time and Procedure Time) can lead to efficient operating room (and even emergency room) decision-making. Identifying the most high-impact metrics and learning how small changes to them can lead to significant improvement in the SSR can help clinical institutions develop low-cost, low-effort strategies to achieve more on-time surgery completions. An example is changing the teaching day, which requires minimal effort and no cost and can have an enormous impact on the SSR.

Our comprehensive descriptive analytics of the data collected from the 4-joint Arthroplasty surgeries at our institution revealed the following insights. Note that these insights are selected from dozens of individual analyses performed on the collected data points.

- If we analyze the three sets of metrics (staff, patient, and time) based on how strongly they influence/impact the SSR, staff metrics take the lead. Patient metrics had the most minimal impact on the SSR.
- Time in Room for the first case of the day influences the SSR of the rest of the cases. Hence it is very important to start the day on time.
- Even though it may seem logical that the most experienced professionals (especially surgeons), with the highest number of surgeries on their record, would complete more procedures on time than their less experienced peers, the analysis revealed that it was not the case. The experience of medical professionals had almost no impact on the SSR, our analytics says it could be their patient selection or their surgical time or both.
- The SSR jumped as high as 45% from the least successful month (14% SSR) to the most successful one (59% SSR).
- The variance among SSR on different days of the week is less significant than months, but it's still significant, i.e., 19%. The highest SSR for a day reached 46%, while the lowest was around 27%.
- On average, male patients required five more minutes per procedure compared to female patients.
- The age and ASA classification of the patients had a significant impact on anesthesia metrics but not on the overall procedure duration.

It's important to understand that many of the above observations are limited by the spread of data which may have influenced the accuracy of some resulting patterns. Most of the outliers are in the extremes. For age, the bulk of the data points is concentrated between the ages of 47 and 76. For BMI, most data points/patients fall between 22.1 and 27.1. As for ASA, most patients are classified as Class II or III, with only a fraction in Class I or IV. Sex is the only variable that's safe from this uneven spread.

DA also helps us identify unique and useful patterns and trends that emerge from the data by combining and comparing a couple of metrics together and studying their relationship (17). However, the effectiveness of DA goes down as dimensionality (i.e., the number of variables/columns of data) increases. It becomes difficult to identify patterns and trends to generate useful insights. Another limitation of DA is the number of variables it can simultaneously handle (18). Some insights can only be generated when more variables are being analyzed at once, and that is where more comprehensive analytical techniques (predictive and prescriptive analytics) and machine learning comes into play (19).

Predictive and prescriptive analytics can take dozens of variables/parameters/dimensions into account and simultaneously analyze them to identify more complex and insightful patterns. Machine learning algorithms are significantly more powerful and can handle thousands of parameters and variables at once. This sophistication allows them to determine patterns and generate insights that DA is unable to generate, though it doesn't undermine its usefulness.

Another reason more sophisticated analytical techniques and machine learning algorithms are prioritized over DA is the depth of analysis. Since it can only handle a few variables and dimensions at once, many of the insights generated are naturally shallow and may simply lead to ineffective or resource-intensive actions taken to achieve a desired outcome. In the worst cases, the conclusions may be wrong and lead to potentially damaging decisions. Using these insights to infer cause and effect without exploring the deeper relationships of these variables to others may lead to wrong conclusions (20).

This also limits the portability of decision-making frameworks based on DA. The decisions and conclusions of a DA may not apply to a different healthcare setting and cannot be generalized for a broader range of scenarios (21). They are usually only valid for the data at hand, and decisions made using the DA can only effectively apply to the source of the data (in this case, the 4-joint surgical OR).

As DA is rooted in the past, it neither informs us about the future nor helps us predict how changes in the current variables will influence the future. This is the domain of predictive analytics, which gives us a glimpse of the future and helps us positively influence it by making relevant changes.

This limitation is tied to the DA's inherent limitation of identifying what happened (patterns) but not why and how it happened. Since it doesn't identify the cause that leads to the apparent effect (pattern/trend), its effectiveness is limited when it comes to decision-making. In contrast, ML algorithms like Decision Tree and Linear Regression that also incorporate DA's strength (explainability) shed a more comprehensive light on the past, and the insights they reveal can be applied to future decision-making (achieving the desired output).

7 Conclusion

The insights generated in our study endorse an important benefit of descriptive analysis, i.e., identifying high-impact metrics. Various analyses can help with the identification of the highest-impact metrics and prevent researchers from assigning more weight to variables/metrics that may seem more impactful than they are due to cognitive bias (like staff experience). In conclusion, the insights can help influence the re-allocation of resources in daily clinical practice to offset inefficiencies in arthroplasty surgeries.

Data availability statement

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

Ethics statement

The Ottawa Health Science Network Research Ethics Board granted an ethics review exemption given that the project was considered to fall under continuous quality improvement rather than human subject research, as per Canada's Tri-Council Policy Statement on Ethical Conduct for Research Involving Humans (22).

The studies involving humans were approved by Ottawa Health Science Network Research Ethics Board (OHSN-REB). The studies were conducted in accordance with the local legislation and institutional requirements. The Ethics Committee/institutional review board waived the requirement of written informed consent for participation from the participants or the participants' legal guardians/next of kin because not required by Ottawa Health Science Network Research Ethics Board (OHSN-REB).

Author contributions

FA, Data curation, Investigation, Methodology, Writing - original draft, Validation, Writing – review and editing. PB, Conceptualization, Methodology, Writing – review and editing. PF, Conceptualization, Methodology, Funding acquisition, Project Administration, Resources, Supervision, Writing – review and editing. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fsurg.2023.1242287/full#supplementary-material

References

1. Aldegheri R, Trivella G, Saleh M. Articulated distraction of the hip conservative surgery for arthritis in young patients. *Clin Orthop Relat Res.* (1994) 301:94–101. doi: 10.1097/00003086

2. Panteli M, Habeeb S, McRoberts J, Porteous MJ. Enhanced care for primary hip arthroplasty: factors affecting length of hospital stay. *Eur J Orthop Surg Traumatolo.* (2014) 24:372–8. doi: 10.1007/s00590-013-1188-z

3. Zhang X, Shi G, Sun X, Zheng W, Lin X, Chen G. Factors influencing the outcomes of artificial hip replacements. *Cells Tissues Organs.* (2019) 206:206. doi: 10.1159/000500518

4. Mistry JB, Chughtai M, Elmallah RK, Le S, Bonutti PM, Delanois RE, et al. What influences how patients rate their hospital after total hip arthroplasty? *J Arthroplasty.* (2016) 31:2422–5. doi: 10.1016/j.arth.2016.03.060

5. Fairley M, Scheinker D, Brandeau ML. Improving the efficiency of the operating room environment with an optimization and machine learning model. *Health Care Manag Sci*. (2019) 22:756–67. doi: 10.1007/s10729-018-9457-3

6. Khalifa M. Health analytics types, functions and levels: a review of literature. *Stud Health Technol Inform*. (2018) 251:137–40. doi: 10.3233/978-1-61499-880-8-137

7. Grimes DA, Schulz KF. Descriptive studies: what they can and cannot do. *Lancet*. (2002) 359:145–9. doi: 10.1016/S0140-6736(02)07373-7

8. Mehta N, Pandit A. Concurrence of big data analytics and healthcare: a systematic review. *Int J Med Inform*. (2018) 114:57–65. doi: 10.1016/j.ijmedinf.2018.03.013

9. Al Zoubi F, Gold R, Poitras S, Kreviazuk C, Brillinger J, Fallavollita P, et al. Artificial intelligence-driven prescriptive model to optimize team efficiency in a high-volume primary arthroplasty practice. *Int Orthop*. (2022) 47:343–50. doi: 10.1007/s00264-022-05475-1

10. Perdue ML, Arnold F, Li S, Donabedian A, Cioce V, Warf T, et al. The future of cell culture-based influenza vaccine production. *Expert Rev Vaccines*. (2011) 10:10. doi: 10.1586/erv.11.82

11. Caban JJ, Gotz D. Visual analytics in healthcare—opportunities and research challenges. *J Am Med Inform Assoc*. (2015) 22:260–2. doi: 10.1093/jamia/ocv006

12. Sampson M, Tetzlaff J, Urquhart C. Precision of healthcare systematic review searches in a cross-sectional sample. *Res Synth Methods*. (2011) 2:119–25. doi: 10.1002/jrsm.42

13. Althobaiti K. Surveillance in next-generation personalized healthcare: science and ethics of data analytics in healthcare. *New Bioethics*. (2021) 27:295–319. doi: 10.1080/20502877.2021.1993055

14. Choong ALC, Shadbolt C, Dowsey MM, Choong PFM. Sex-based differences in the outcomes of total hip and knee arthroplasty: a narrative review. *ANZ J Surg*. (2021) 91:553–7. doi: 10.1111/ans.16299

15. Buljac-Samardzic M, Doekhie KD, Van Wijngaarden JDH. Interventions to improve team effectiveness within health care: a systematic review of the past decade. *Hum Resour Health*. (2020) 18. doi: 10.1186/s12960-019-0411-3

16. Beck MJ, Hensher DA. Australia 6 Months after COVID-19 restrictions part 2: the impact of working from home. *Transp Policy (Oxf)*. (2022) 128:274–85. doi: 10.1016/j.tranpol.2021.06.005

17. Deshpande PS, Sharma SC, Peddoju SK Predictive and prescriptive analytics in Big-data era. *Security and data storage aspect in cloud computing. Studies in big data*. Singapore: Springer (2019). p. 978–81. doi: 10.1007/978-981-13-6089-3_5

18. Hund M, Böhm D, Sturm W, Sedlmair M, Schreck T, Ullrich T, et al. Visual analytics for concept exploration in subspaces of patient groups. *Brain Inform*. (2016) 3:233–47. doi: 10.1007/s40708-016-0043-5

19. Oo MCM, Thein T. An efficient predictive analytics system for high dimensional big data. *J King Saud University—Comput Inf Sci*. (2022) 34:1521–32. doi: 10.1016/j.jksuci.2019.09.001

20. de Mast J, Steiner SH, Nuijten WPM, Kapitan D. Analytical problem solving based on causal, correlational and deductive models. *Am Stat*. (2023) 77:51–61. doi: 10.1080/00031305.2021.2023633

21. Polit DF, Beck CT. Generalization in quantitative and qualitative research: myths and strategies. *Int J Nurs Stud*. (2010) 47:1451–8. doi: 10.1016/j.ijnurstu.2010.06.004

22. Khaliq Y. Tri-council policy statement: ethical conduct for research involving humans. *Encyclopedia of clinical pharmacy*. Panel on Research Ethics (2002).

Supplementary Material

Factors Influencing Delays and Overtime During Surgery: A Descriptive Analytics for High Volume Arthroplasty Procedures

Farid Al Zoubi*, Paul Beaule, Pascal Fallavollita

* Correspondence: Farid Al Zoubi, falzo100@uottawa.ca

1 Supplementary Figures

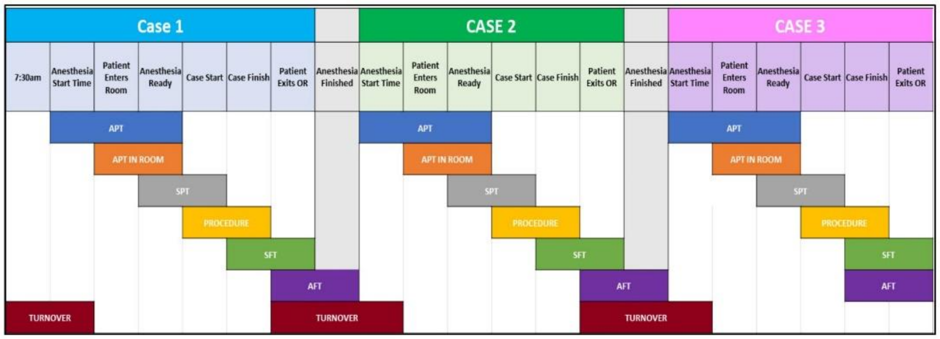


Figure 1: Surgical workflow steps [9]

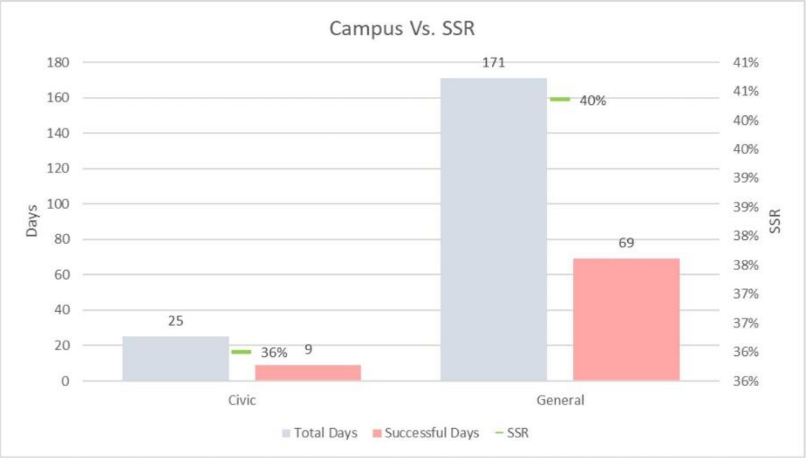


Figure 2: Average SSR for both campuses is 39%.

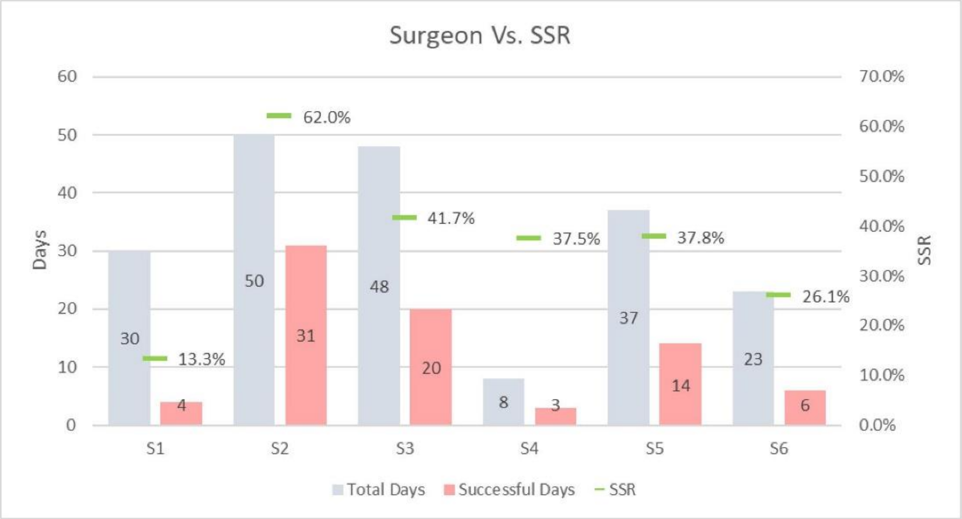


Figure 3: Surgeon's experience did not influence SSR.

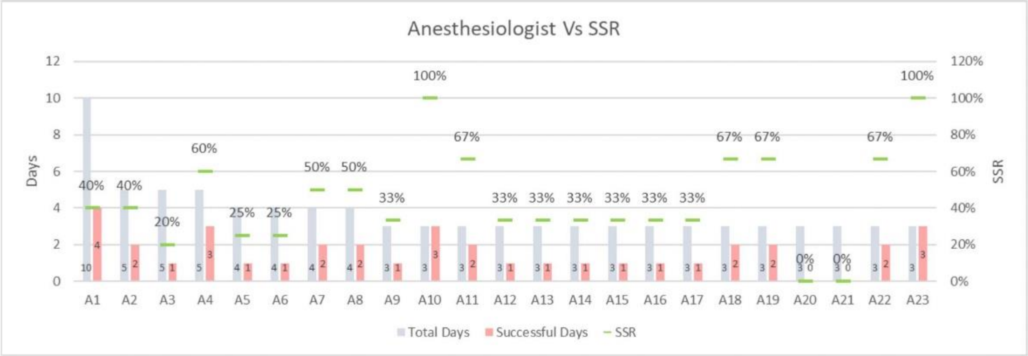


Figure 4: Anesthesiologist's experience did not influence SSR.

Figure 5: Circulating nurse's experience did not influence SSR.

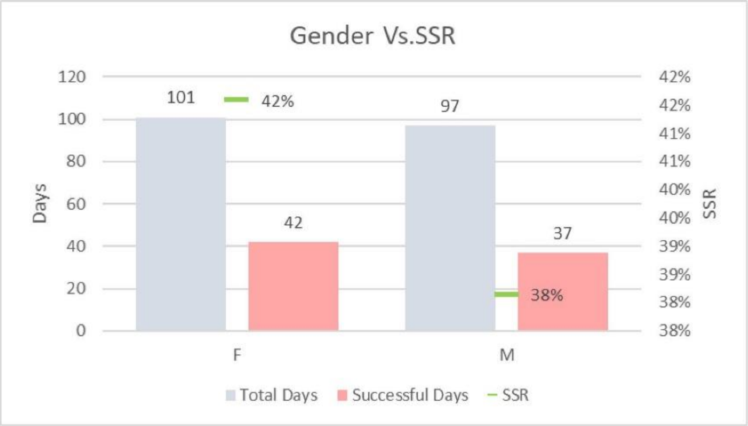


Figure 6: four percent difference in SSR between male and female patients.

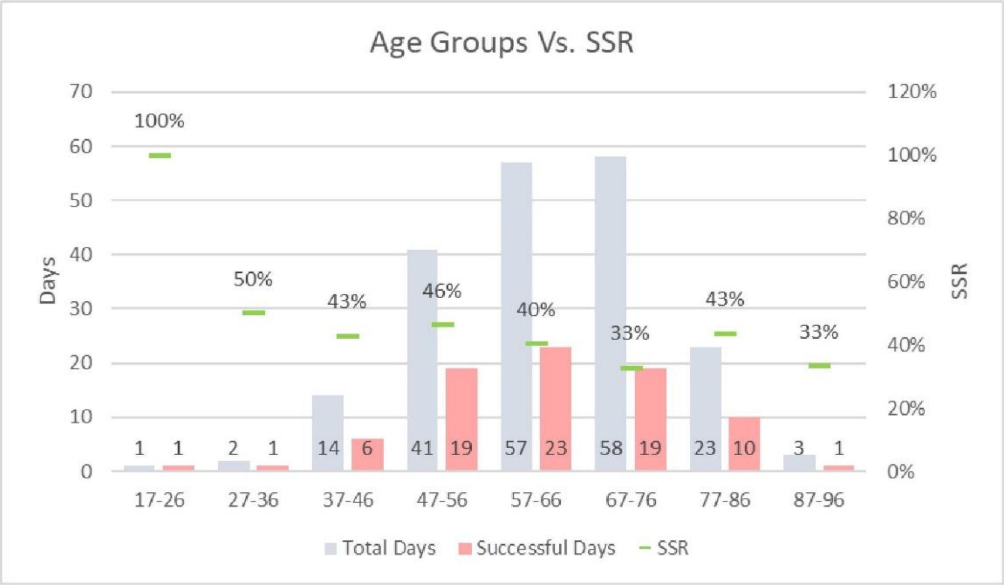


Figure 7: Differences in SSR between patient age groups.

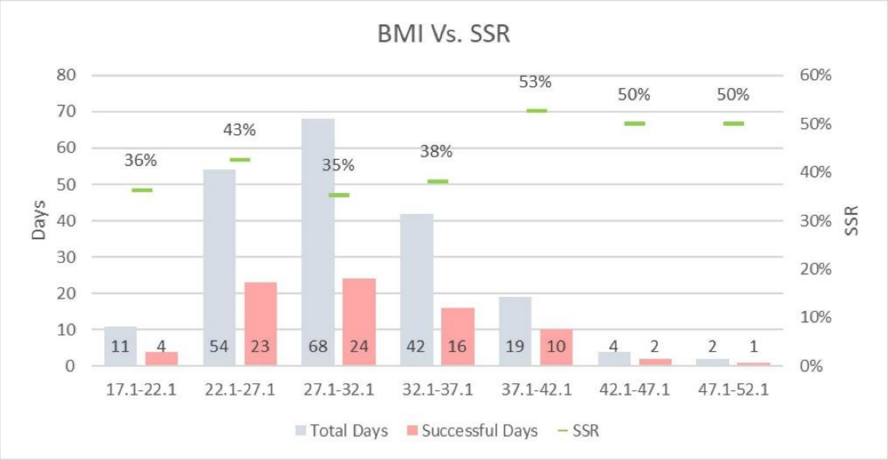


Figure 8: BMI and SSR did not follow same pattern.

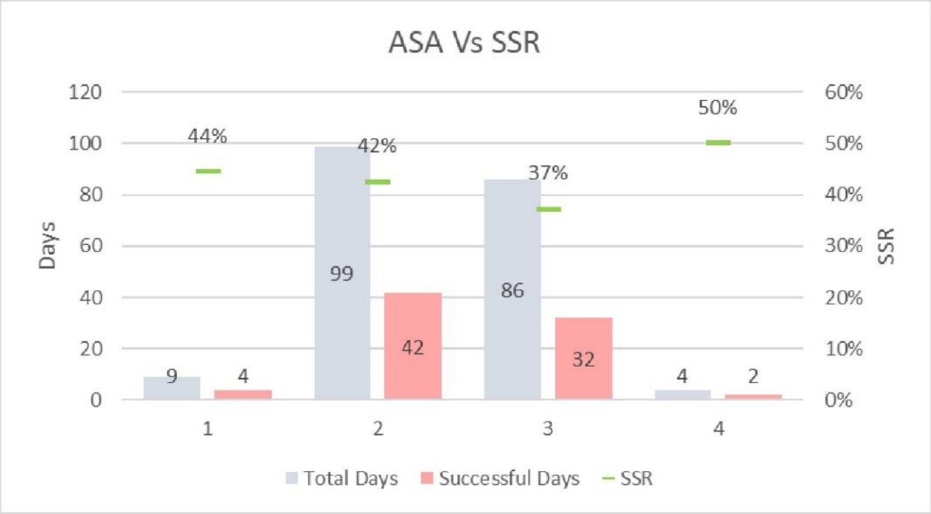


Figure 9: SSR is higher for a safer ASA class and lower for a riskier class.

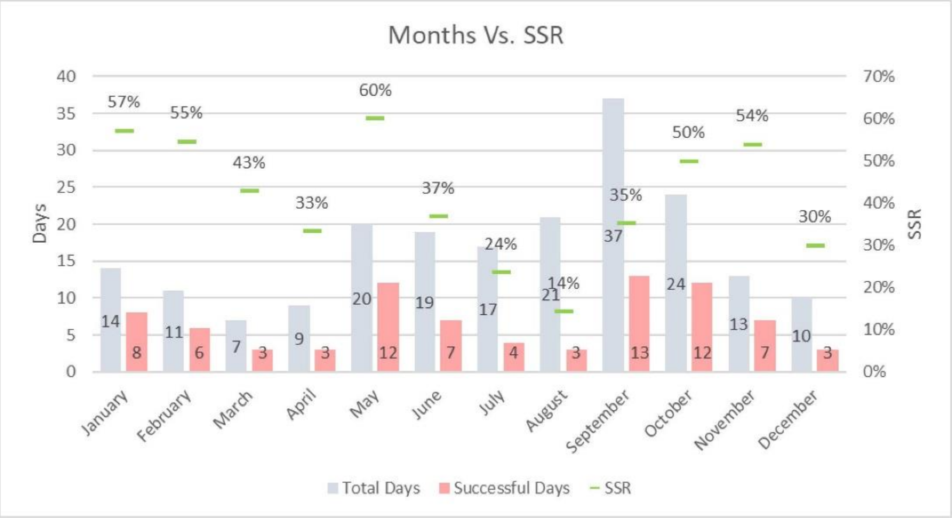


Figure 10: The month of May has the highest SSR opposite to the month of August.

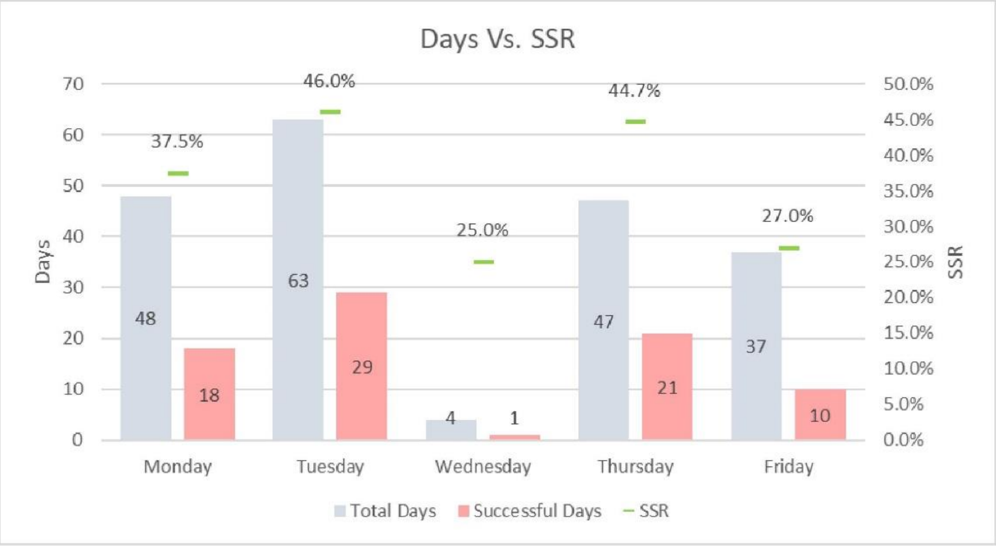
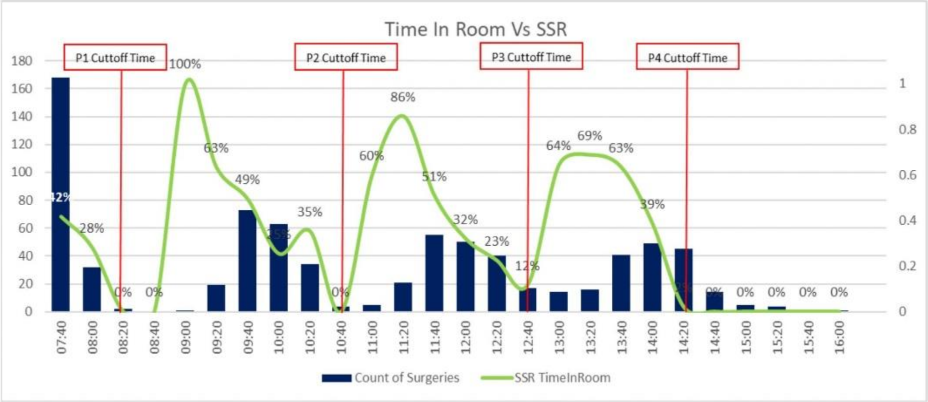


Figure 11: Days after and before weekends have the worst SSR.



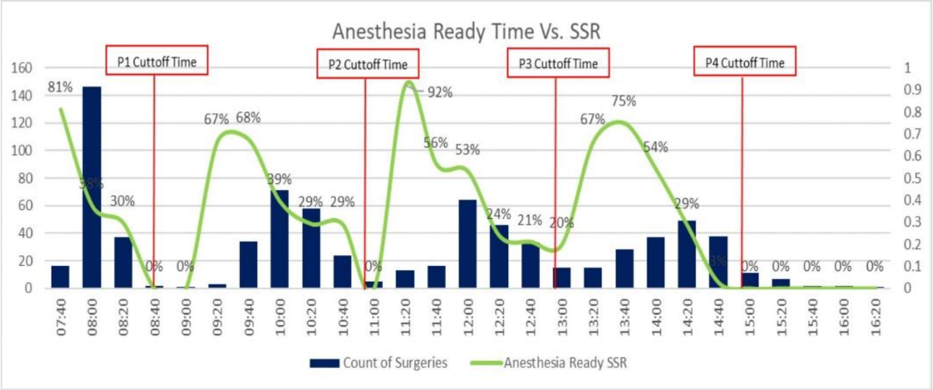
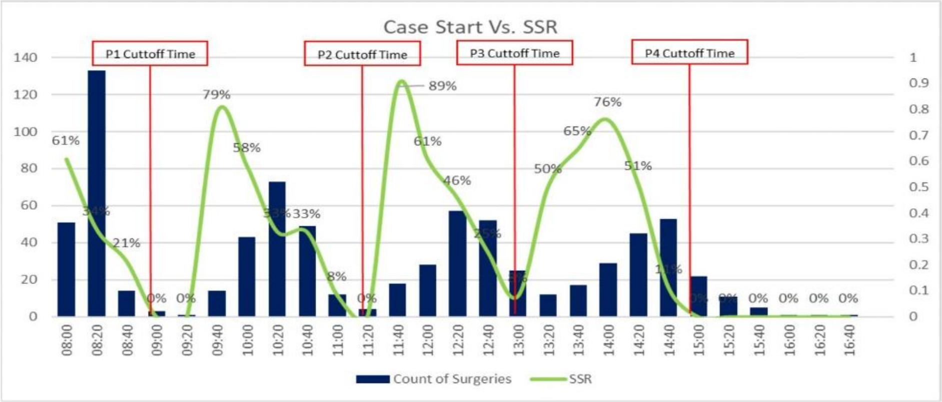


Figure 12: The SSR is higher for surgeries in which the patient was in the room and anesthetized close to the scheduled time.



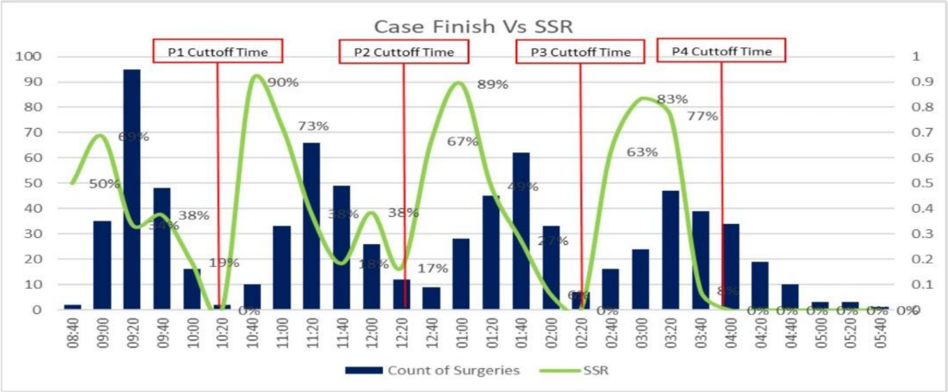


Figure 13: Procedures that started at the allotted time slots had a much higher SSR rate

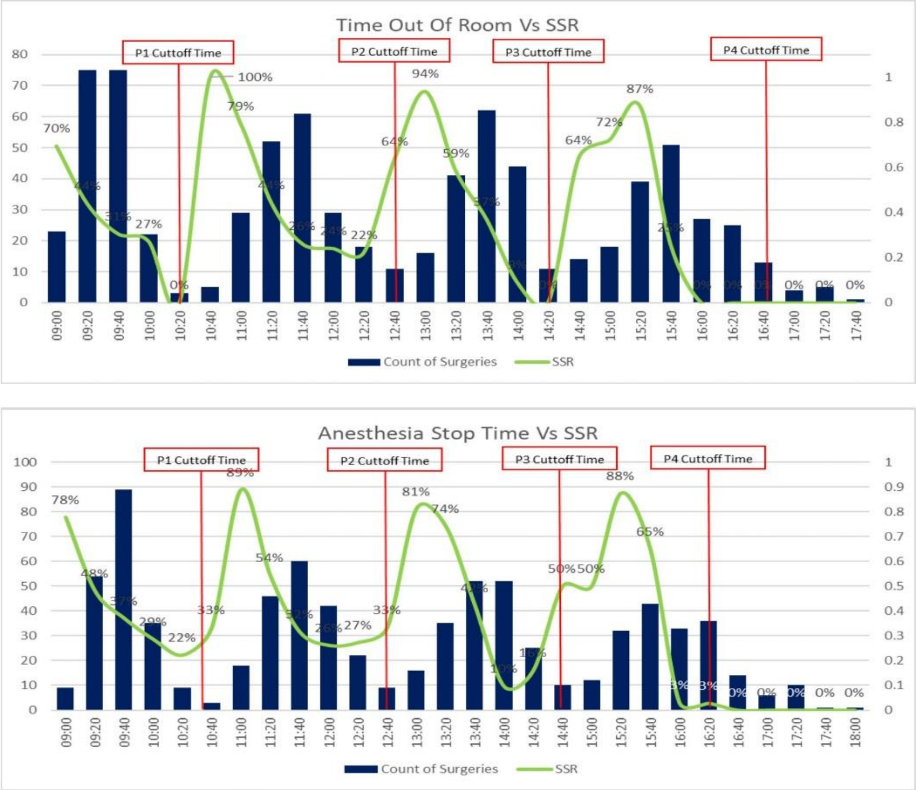


Figure 14: The earlier the (out of room/anesthesia stop) times the higher SSR.

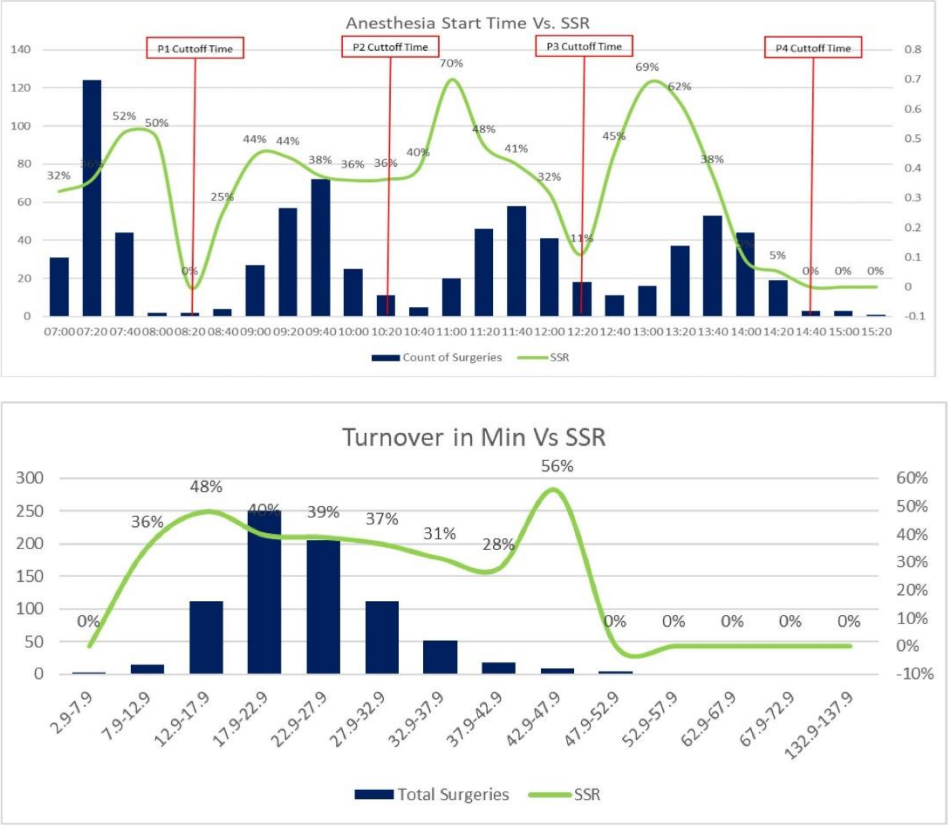


Figure 15: Different patterns are observed for turnover (in minutes) and anesthesia start time.

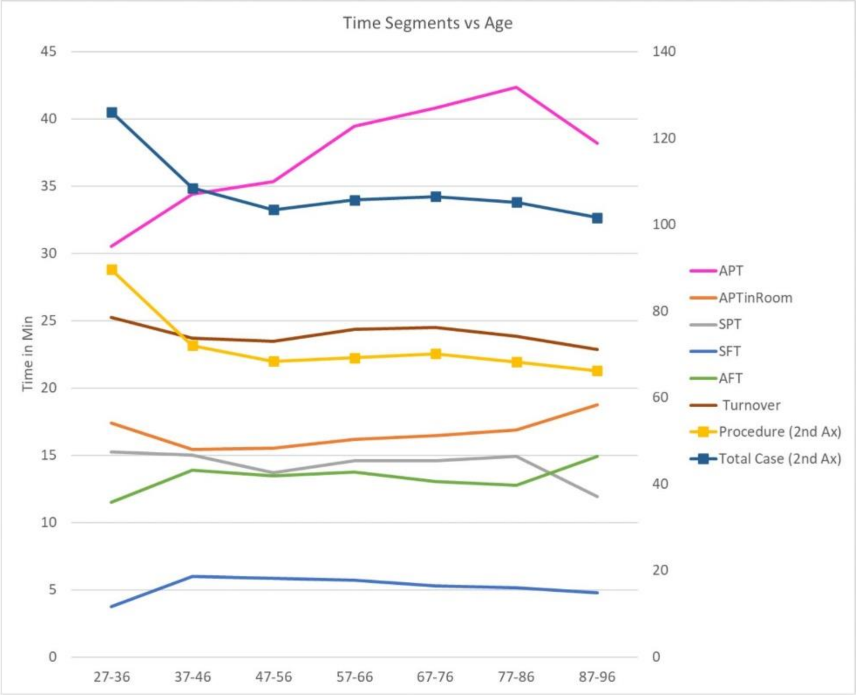


Figure 16: Anesthesia preparation and in-room time go up with age.

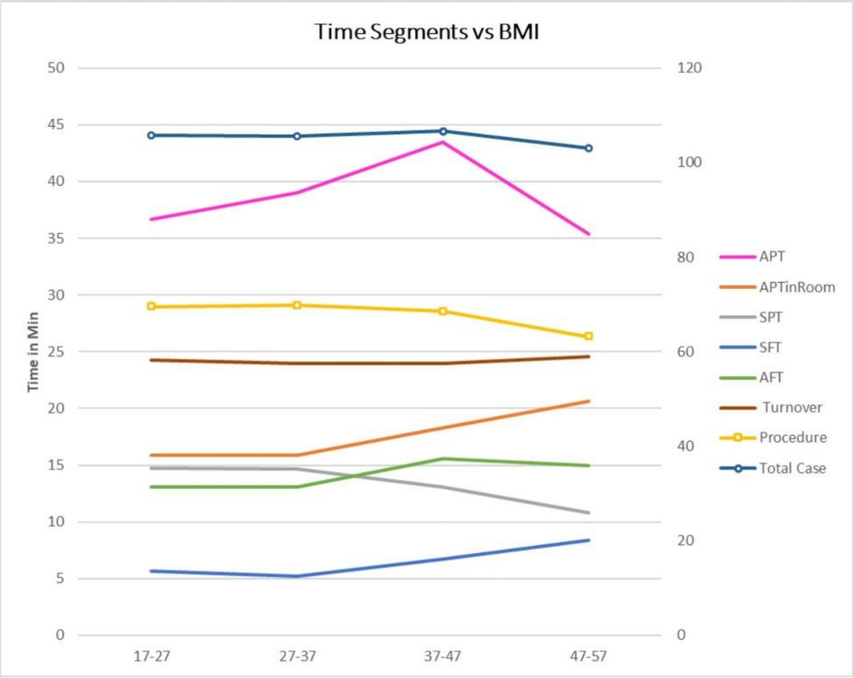


Figure 17: Relationship between BMI and time metrics for surgery.

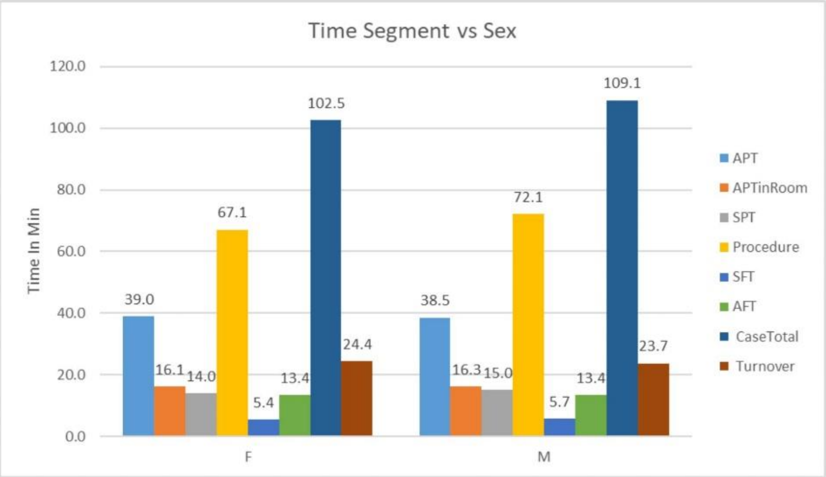


Figure 18: Male patient surgeries are on average 5 minutes longer than those for female patients.

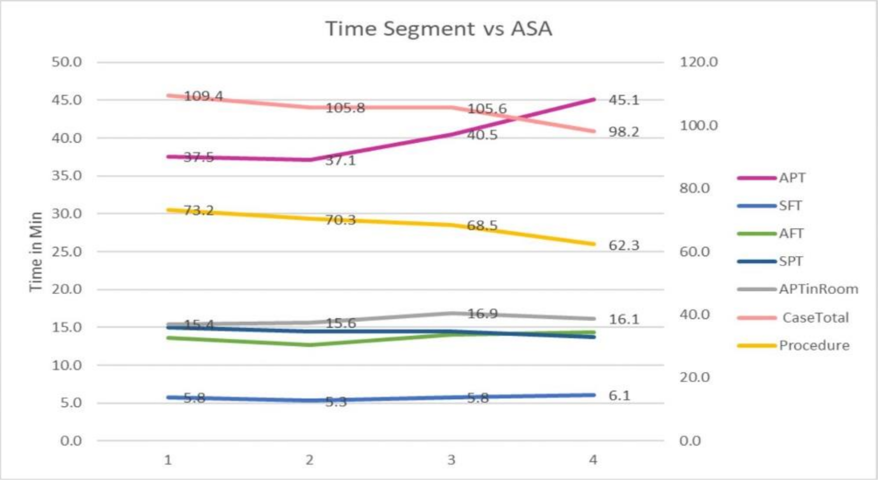


Figure 19: Only APT is directly positively correlated to ASA.

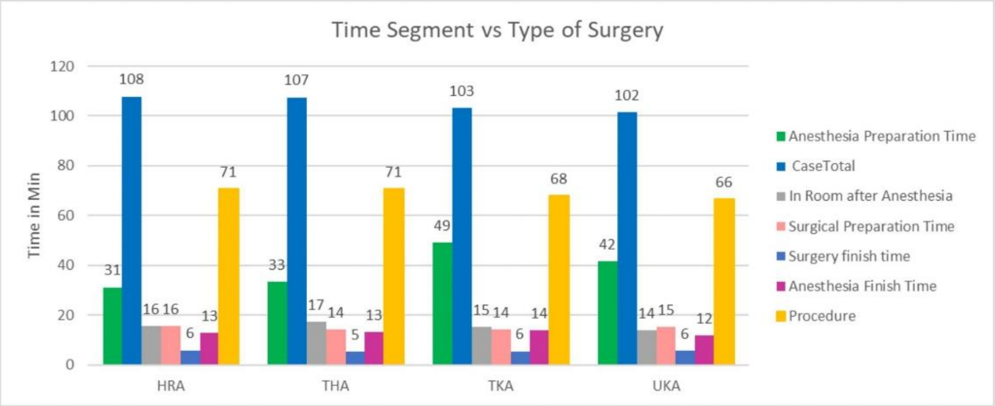


Figure 20: Variance in time is not significant as it is for SSR for different types of surgeries.

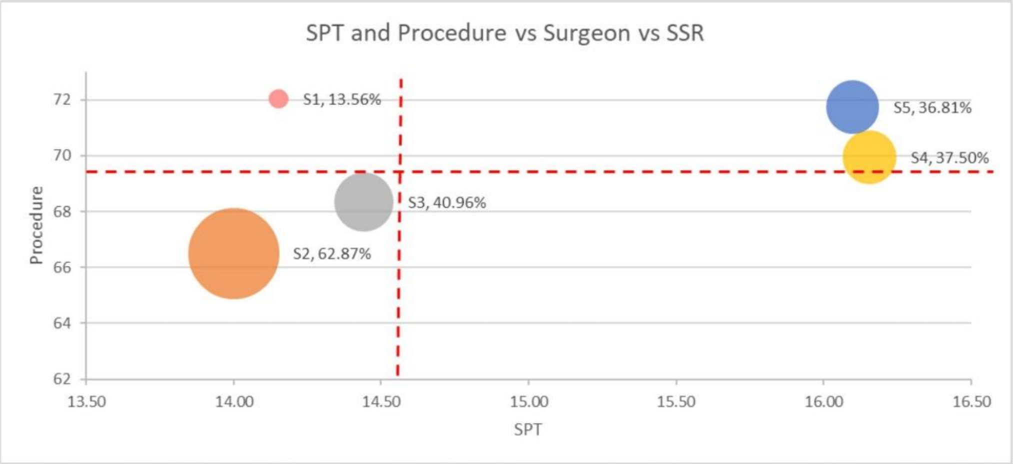


Figure 21: Surgeons, in general, with high SSR have lower SPT and procedure time. The size of the sphere represents the SSR of the surgeon.

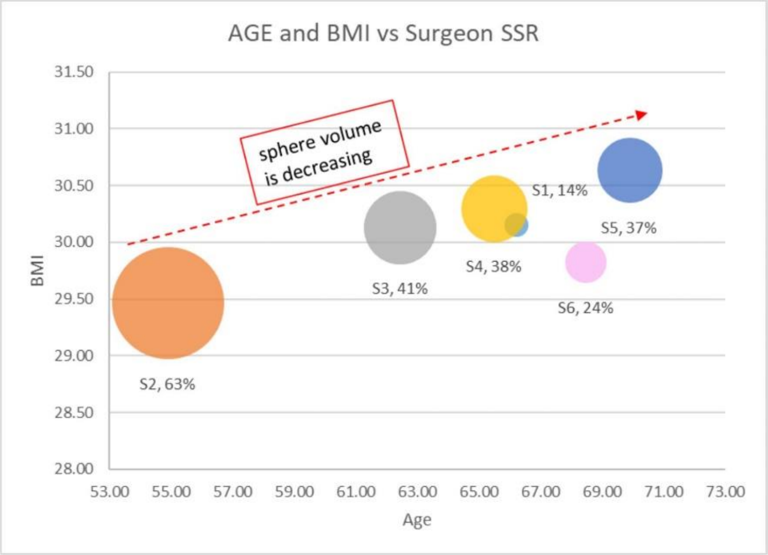


Figure 22: Age is more important on surgeon SSR than BMI.

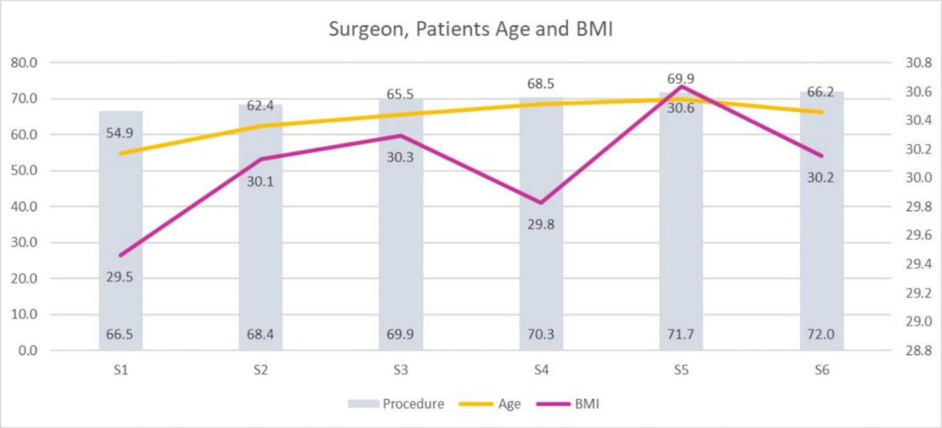


Figure 23: Surgeons with higher SSR have operated on patients with lower BMI and age.

2 Supplementary Tables

Time Metrics	Staff (Team) Metrics	Patient Metrics	Safety metrics
Anesthesia Preparation time	Surgeon	Campus	90-day Readmissions
Anesthesia Preparation time in Room	Anesthesiologist	Type of Surgery	Reason for Readmission
Anesthesia Start time	Circulator Nurse 1	Type of Anesthesia	Length of Stay
Time in Room	Circulator Nurse 2	Sex	
Anesthesia Ready time		Age	
Anesthesia Stop Time		BMI	
Anesthesia Finish Time		ASA	
Surgical Preparation Time			
Case Start			
Case Finish			
Surgery finish time			
Turnover			
Surgery (Procedure) Time			
Time Out of Room			
Case no			
Date			

Table 1: Metrics and their categories

Metric	Pearson Correlation With SSR	Spearman Correlation With SSR
Age	-0.088	-0.092
ASA	-0.038	-0.046
Turnover	-0.079	-0.069
APT	-0.181	-0.195
AFT	-0.067	-0.085
APTinRoom	-0.094	-0.109
BMI	0.018	0.004
SFT	-0.107	-0.124
SPT	-0.106	-0.109
Procedure	-0.150	-0.197
Sex	-0.031	-0.031

Table 2: Correlation of each numerical metric to the SSR

Chapter 3: Frameworks for AI-based Surgical Transformation (FAST)

3.1. Summary

This journal focuses on categorizing the data from the previous study into three distinct categories: patient-related metrics, time-related metrics, and staff-related metrics. For each category, a machine-learning-based framework was developed to predict the SSR. The frameworks are then compared to evaluate their effectiveness. These frameworks, which are flexible and generalizable, form an integral part of the predictive analytics component of my thesis work. Their analysis and evaluation contribute to a deeper understanding of their predictive capabilities and their potential application in optimizing surgical outcomes.

3.2. Methodology

Solely developing ML models leads to an unhealthy accumulation of technical, cultural, and research debt, which elevates the AI-framework's risk profile. Source code is just one of many other elementary building blocks of an ML solutions. From a lean perspective, value to the customer can only be assessed for the entire AI-framework. Therefore, we must stop thinking about ML models in isolation and expand our vision to conceptualize their role as the core of a dynamic, constantly adapting AI-framework.

In this chapter, we propose thorough frameworks that can be transformed into decision support systems. These frameworks are crafted to be flexible and adaptable, making them suitable for repurposing across a range of surgeries and healthcare settings, including those markedly different from arthroplasty procedures. Through a comparative analysis of these frameworks, our aim is to aid various healthcare institutions and professionals in pinpointing the most fitting framework for their optimization requirements.

However, **the central theme of this thesis** revolves around **enhancing arthroplasty OR throughput** specifically by **optimizing team efficiency**. In our case, the chosen framework for achieving this improvement is WTMF, as detailed in the submitted article below. The ML output all the frameworks were discussed for the sake of comparison and as a suggestion of how can we in the future benefit from the spared data, i.e. data that was not used to develop the WTMF and the PAS.

Within Appendix B, I've incorporated further details, drawing from both the published article and additional content not covered in the submitted version. These additions elaborate on the distinctions between the input and output of both machine learning models and frameworks, providing additional examples, details, and technical insights to showcase these models and frameworks.

3.3. Fundamental research contributions

Three framework proposals were introduced, each customized for a specific type of suitable data. The examination and comparison of the machine learning module's output were specifically concentrated on each framework. Notably, at this stage, no details were provided about the recommendation system or the other components for any of the frameworks.

3.4. Author Contributions

Author	**Contribution to the study**
Farid Al Zoubi	Data curation, Investigation, Methodology, Algorithm design and evaluation, Writing - original draft, Validation, Writing – review and editing
Paul Beaulé	Clinical translation concepts; proof reading
Pascal Fallavollita	AI-translation concepts, proof reading

3.5. The Article

The following article was submitted to frontiers in Artificial Intelligence Machines and Public Health.

Frameworks for AI-based Surgical Transformation (FAST)

Farid Al Zoubi, PhD[1], falzo100@uottawa.ca

Paul E. Beaulé, MD FRCSC[2], pbeaule@toh.ca

Pascal Fallavollita, BEng, PhD[3], pfallavo@uottawa.ca

Institutional Affiliations and Addresses:

[1]School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, Ontario, Canada

[2]Division of Orthopaedic Surgery, The Ottawa Hospital, Ottawa, Ontario, Canada

[3]Interdisciplinary School of Health Sciences, University of Ottawa, Ottawa, Ontario, Canada

Keywords: Machine Learning, Healthcare delivery, Predictive analytics, Prescriptive analytics.

Abstract

Purpose: The North American healthcare system is stretched thin. With an aging population, the system is experiencing a powerful surge in demand. Between 2014 and 2019, arthroplasty surgeries in Canada experienced a steady increase with 20.1%. We believe that a lack of an efficient approach, long-term sustainable vision, and misallocation of appropriate resources undermined the initiative's potential. Disorganized and suboptimal use of OR time and space alone inflate surgical costs by 30%. An in-depth understanding of how all the critical factors (i.e., patients, surgical team members, and OR workflows) fit together can help identify bottlenecks and improve efficiency.

Methods: In this paper, we present three AI-driven prescriptive frameworks that aim to do more than just forecast the surgical success rate (SSR), a metric directly influencing OR throughput, i.e., completing 4-arthroplasty in 8-hr window. These frameworks also assist in creating the necessary combination of either patients, surgical teams, or a set of benchmarks required to attain specific SSR goals.

Results: Three frameworks have been created, each featuring its own Machine Learning model: The Patient Scheduling Framework (PSF), The Surgical Team Scheduling Framework (STSF), and The Workflow/Time Monitoring Framework (WTMF). They entail the collaboration of all key stakeholders, including patients' demographics, surgical professionals, and time management. Furthermore, they encompass all stages of the surgical process, including pre-operation, intra-operation, and post-operation phases

Conclusion: The choice for the right ML model is governed by far more than just their performance numbers. Multiple factors, from the difficulty of obtaining relevant inputs to the potency of the optimization insights they helped generate, were considered. The comprehensive nature of the frameworks is both flexible and generalizable and allows them to be repurposed for a wide range of surgeries and healthcare environments, even the ones that are radically different from arthroplasty procedures.

1 Introduction

The North American healthcare system is stretched thin. With an aging population, the system is experiencing a powerful surge in demand while delivery is being throttled by regulatory challenges and complex budgets. The bundled care payments are only aggravating the problem ("Canadian Institute for Health Information. (2020) Hip and Knee Replacements in Canada: CJRR Quick Stats, 2018–2019. Ottawa, ON," n.d.). At a 43% productivity level of the healthcare system, the situation is not sustainable (Fairley, Scheinker, and Brandeau 2019). Between 2014 and 2019, arthroplasty surgeries in Canada experienced a steady increase with 20.1% (hip replacement) and 22.5% (knee replacement), respectively (Cram et al. 2018). In order to meet the increasing needs of the population and cut down the wait time to target levels, multiple initiatives were taken. These included dedicated funds (federally regulated) for quality-based procedures, creating four joint operating room (OR) days to increase the number of surgeries performed, and overlapping swing rooms (Waly et al. 2020)(Beaulé, Frombach, and Ryu 2015). Despite being a promising initiative, four joint ORs in our institution exhibited a discouraging 39% Surgical Success Rate (SSR), the ability to complete 4-joint replacements in 8-hr window; signifying that close to 60% of the four joint replacement surgery days went into overtime (Richard Gold 2022). We believe that a lack of an efficient approach, long-term sustainable vision, and misallocation of appropriate resources undermined the initiative's potential. Disorganized and suboptimal use of OR time and space alone inflate surgical costs by 30% (Fairley, Scheinker, and Brandeau 2019). An in-depth understanding of how all the critical factors (i.e., patients, surgical team members, and OR workflows) fit together can help identify bottlenecks and improve efficiency.

Existing solutions that facilitate and optimize hospital efficiency are classified as either clinical or non-clinical approaches. Clinical approaches refer to the most commonly recognized roles in the industry. These encompass medical practitioners who offer direct patient care and typically dedicate numerous years to rigorous education and specialization in a particular medical field. Conversely, non-clinical approaches refer to roles that do not involve any form of medical treatment or testing (Shohreh Majd; Zohreh Majd 2023). They can be categorized into three facets: (i) introduction of new assets to improve hospitals' efficiency (C. Yeoh 2018); (ii) leverage of rhetorical data and descriptive analytics (data driven) to change, evaluate, or rearrange existing assets in a hospital, i.e., no machine learning involved (Porta et al. 2013); and (iii) data driven solutions leveraging machine learning (ML) torniquets (prescriptive and predictive analytics). When specifically utilizing machine as shown in Figure 1, the efforts to improve efficiency can be directed towards scheduling patient surgeries, scheduling surgical team/resources, and optimising surgical workflow time. The remainder of this section presents a literature review concerning these endeavors.

1.1 Patient scheduling allocation

The work in the field of patient scheduling can be categorized based on the type of patient data utilized for modeling. Authors in (Eshghali et al. 2023)focused on geographical data, such as the distance of patients from the hospital and the potential impact of traffic on assessing the risk of no-shows or associated delays. They used this information to schedule patients and allocate operative days accordingly. In contrast, authors in (Samorani and Blount 2020) centered their work around sociodemographic data to extract patient characteristics. The scheduling data included factors like a patient's history of previous no-shows, the number of past appointments, the lead time for scheduling appointments, and whether the patient had insurance. Their machine learning scheduler predicted the likelihood of a patient not showing up and determined how many additional patients to schedule for a day to prevent the OR from being unused in case of cancellations. Another type of data found in the literature related to patient scheduling is time-related metrics that could influence patient decisions and delays. This category is exemplified in (Salah and Srinivas 2022), where the authors considered metrics like patient waiting time, doctor idle time, and overtime. They proposed a system that predicts the likelihood of no-shows and then schedules appointments based on this risk assessment.

1.2 Surgical team allocation

When it comes to surgical staff allocation, there has been a substantial amount of work in utilizing machine learning (ML) for Human Resource (HR) allocation in various industrial sectors. However, the application of ML in healthcare has been relatively limited (Garg et al. 2022). The predominant approach in HR allocation has involved the use of unsupervised ML techniques, particularly clustering. These efforts primarily focus on grouping individuals with similar characteristics, such as grouping patients with specific diseases (Myszczynska et al. 2020). Alternatively, some initiatives aim to create diverse groups with a wide range of capabilities, as seen in the case of study groups for online education (Maina, Oboko, and Waiganjo 2017). In clinics, a significant portion of HR allocation efforts has been dedicated to matching the right physician with the appropriate patients. This is done to address the demands of patients, optimize resource utilization, and ensure the delivery of high-quality healthcare in settings with limited resources (Rudra Kumar, Pathak, and Gunjan 2022; Lazebnik 2023). Additionally, there have been instances where ML has been employed to allocate teams and working hours during high-traffic periods, thereby preventing congestion, as demonstrated in (Dehnoei et al. 2022).

1.3 Workflow / Time allocation

Much of the workflow time-optimisation solutions aim to predict a certain aspect of an operation; this can be predicting an event before surgery (Maheshwari et al. 2017), after surgery (Fairley, Scheinker, and Brandeau 2019), (Schiele, Koperna, and Brunner 2021), or even the duration of the surgery itself (Bartek et al. 2019). Concerning pre-surgical events, work usually attempted to develop a model that could successfully predict anesthesia preparation time. Instead for post-surgical events, authors addressed their hospital's post-anesthesia care unit (PACU) crowding and predicting each patient's PACU stay time based on their surgery. Alternatively, authors built a ML model that predicted the intensive care unit's occupancy level in terms of number of beds. Other solutions aim to establish benchmarks that help maximize efficiency or achieve set goals (Erekat et al. 2020).

1.4 Contributions

In this paper, we present three AI-driven prescriptive frameworks that aim to do more than just forecast the SSR, a metric directly influencing OR throughput. These frameworks also assist in creating the necessary combination of either patients, surgical teams, or a set of benchmarks required to attain specific SSR goals.

First, we are proposing a new approach to boost OR throughput through patient scheduling by incorporating demographic patient data, which we believe is the first of its kind. Our scheduling system recommends which groups of patients should be assigned to each OR to ensure no delays or overtime hours. Second, we are proposing to optimize the surgical team allocations by identifying which of the surgeons, nurses, and anesthesiologists are grouped together for a specific surgery to ensure the best possible OR throughput. Third, in terms of workflow and time allocation, unlike previous literature, we are considering all phases of surgery, including pre-operation, intra-operation, and post-operation, rather than focusing on just one phase. Our proposed frameworks rely on machine learning models and have been validated using data from 4796 patient arthroplasty surgeries.

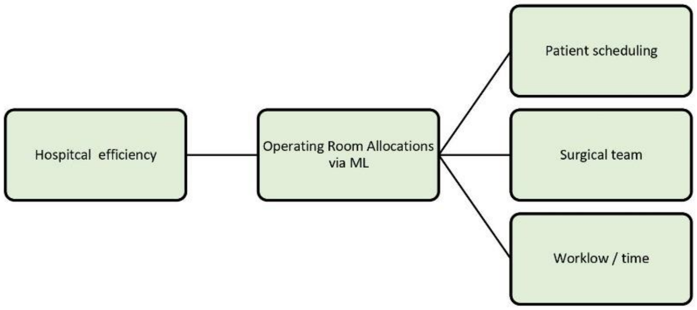


Figure 1: Overview of ML endeavors to optimise operating rooms efficiency

2 AI-driven Frameworks & ML Models: Components and Complexity

A ML model is the core of a functional AI product, not the entirety of the package. The model must be augmented, sustained, and connected with a wide range of functionalities to reach its full potential and offer value to the customers. Collectively, the model and all its functionalities it relies upon to offer tangible value are called an AI-driven system. In this paper we refer to it as AI-frameworks (Jordan and Mitchell 2015). One negative trend that has emerged over the years is that the primary focus is the creation of the ML model to the detriment of all the other aspects of the AI-framework. This phenomenon is referred as model-centric bias, i.e., perceiving the whole AI-framework from the ML model's perspective (Sculley et al. 2015). This narrow approach naturally leads to a range of negative consequences for the AI-framework.

Concentrating solely on developing the ML models leads to an unhealthy accumulation of technical, cultural, and research debt, which elevates the AI-framework's risk profile. Source code is just one of the two elementary building blocks of an ML model - the other is data (Jain et al. 2020). High-quality training and testing data are crucial for the success of an ML model and, consequently, the AI product. From a lean perspective, value to the customer can only be assessed for the entire AI-framework. Therefore, we must stop thinking about ML models in isolation and expand our vision to conceptualize their role as the core of a dynamic, constantly adapting AI-framework.

It's crucial to understand that the bulk of an AI-framework's complexity resides outside the model source code (Sculley et al. 2015). An AI-framework's technical debt is disproportionately high compared to a typical software since it's a combination of code-related and ML-specific debt. Complexity is not a focal point of this work but we will shed some light on typical complexity aspects of an AI framework, data collection and feature extraction, to give an appreciation of the scope of the problem it presents.

3 Arthroplasty surgeries: from pre-op to post-op

For elective joint replacement surgery, patients are referred either through their primary care physician or a central intake clinic for hip and knee arthritis where triage is done by an advanced physiotherapist. This is to ensure that the patient has completed a course of non surgical management for their arthritic pain which would include such things as anti-inflammatories, physical therapy. After the orthopedic surgeon in consultation and joint replacement surgery has been consented for, the patient is placed on the surgeon's wait list that is organized based on time of consenting as well as priority level based on degree/severity of pain and disability varying from 6 weeks to up to 6-8 months.

The surgeon is allocated a number of surgery days per month to perform surgeries within the public hospital which ultimately dictates the wait time for the patient's surgery. The surgeon's office calls patient and books the surgery approximately 6-8 weeks ahead of time so that the patient can prepare for surgery and ensure that are medically fit to undergo the joint replacement. There is an anesthesia assessment 2 weeks ahead of the surgery as well as an education session with a physiotherapist to prepare their home and their recovery in the pre-assessment unit. The vast majority of patients are now done as same day discharge. The patient arrives to the hospital 2hrs ahead of their surgery time in the surgical day care unit (SDCU). They receive pre-operative pain medications as well as antibiotics. They are then brought to the OR suite to undergo surgery under anesthesia. After surgery is completed, they are brought to the post anesthesia care unit (PACU) to be monitored for 1-2hrs i.e., pain under control, alert and oriented, anesthetic medication have worn off and then transferred back to surgical day care unit in order to be mobilized with physiotherapy and ensure they are comfortable and can transfer independently. Nurses ensure with the family present that their pain is under control and that they can urinate on their own prior to being discharged home. Further instructions are provided in regards to self-monitoring for complications as well as when to follow-up with their surgeon.

In terms of surgical team selection, circulator "nurses" in the SDCU, OR suites and PACU are trained and specialized for their areas of work. In other words, a nurse working in SDCU is not necessarily trained/qualify to work in PACU. Nurses in the OR suite require specialized training that take several months. Again, a nurse in PACU cannot simply work in the OR suites. The anesthesiologist are randomly assigned to the OR suite a day or 2 ahead of the surgery day and are not necessarily the same anesthesiologist that assessed the patient in the PACU.

4 Patient Data Set 4796 arthroplasty surgeries

The data is collected from four joint ORs which is defined as a scheduled eight-hour day (7:30am–3:30 pm) where four unilateral joint replacements are performed by the same surgeon (Al Zoubi et al. 2022). A surgical day going having delays and moving into overtime (i.e. past 3:30pm) negatively affects the SSR. Shown in Table 1 are the patient-specific surgery demographics and statistics. We used the same data as in (Al Zoubi et al. 2022) that was collected from the Surgical Information Management System (SIMS) at our institution where time intervals used are a modified version of those defined by the Association of Anesthesia Clinical Directors (AACD): anaesthesia preparation time (APT); patient in-room to anaesthesia ready, surgical preparation time (SPT); anaesthesia ready to procedure start, procedure; procedure start to procedure finish, anaesthesia finish time (AFT); procedure finish to patient out of room, and turnover time; and first patient exits to subsequent patient in room. APT immediately follows turnover. The proposed AI-driven frameworks required the use of 29 metrics (both quantitative and qualitative) divided into four categories, as shown in Table 2:

1. Patient metrics: seven patient and condition-specific metrics including the type of surgery, BMI, age, etc.
2. Staff (Team) metrics: four individuals, including surgeon, anesthesiologist, and two circulator nurses.
3. Time metrics: fifteen time and identification metrics that include five anesthesia-related metrics, four surgical time metrics, turnover time, etc. The time metrics can help identify the specific inefficiencies and bottlenecks.
4. Safety metrics: Three metrics including 90-day readmission and reasons behind readmission. A readmission tied to the original surgery records were excluded. The three metrics are quality indicators and are not used as input to our frameworks. They assure that the patient safety is not compromised.

Surgeons	**5**	Females	**2461**
Circulating nurses	**44**	Males	**2335**
Anesthesiologists	**152**	Average age	**64.1**
Four joint days	**1199**	Age Range	**17–99**
Cases	**4796**	Average BMI	**29.93**
Total hip arthroplasties (THA)	**1461**	BMI Range	**17.1–51.4**
Total knee arthroplasties (TKA)	**1496**	Average ASA	**2.45**
Hip resurfacing (HR)	**652**	ASA Range	**1-4**
Unicompartmental knee arthroplasties (UKA)	**242**		
Other Procedures (combination)	**945**		

Table 1: Patient-specific surgery demographics and statistics.

Time Metrics	Staff (Team) Metrics	Patient Metrics	Safety metrics
Anesthesia Preparation time	Surgeon	Campus	90-day Readmissions
Anesthesia Start time	Anesthesiologist	Type of Surgery	Reason for Readmission
Time in Room	Circulator Nurse 1	Type of Anesthesia	Length of Stay
Anesthesia Ready time	Circulator Nurse 2	Sex	
Anesthesia Stop Time		Age	
Anesthesia Finish Time		BMI	
Surgical Preparation Time		ASA	
Case Start			
Case Finish			
Surgery finish time			
Turnover			
Surgery (Procedure) Time			
Time Out of Room			
Case no			
Date			

Table 2: Four categories of data input to the AI-driven frameworks.

5 Methodology

5.1 Frameworks for AI-based Surgical Transformation (FAST)

The comprehensive nature of the frameworks is both flexible and generalizable and allows them to be repurposed for a wide range of surgeries and healthcare environments, even the ones that are radically different from arthroplasty procedures. Through a comparison of the decision support frameworks, we aim to help a wide range of healthcare institutions and professionals determine which framework might be ideal for their optimization. A summary is shown in Table 3.

Framework	Advantage	Drawback	Implementation Complexity	Input Feature	Data Type
Workflow/ Time Monitoring	Boost overall team efficiency	Teamwork high dependency	Hard	6	Numerical
	Team self-evaluation	Cost of building the system			
	Real-time output, can improve on spot				
Surgical Team Scheduling	Best efficiency (SSR)	Team polarization	Hard	4	Categorical
	Minimal data and input required	Team Resistance			
	Easy to obtain input	Weaker teams stay stagnant			
Patient Scheduling	Completely independent on other systems	Least efficiency (SSR)	Easy	5	Mix
	No impact nor dependency on OR aspects	Depends on the Surgeon only			

Table 3: A comparison of the proposed frameworks

5.1.1 Patient Scheduling Framework

The Patient Scheduling Framework (PSF), Figure 2, focuses on scheduling patients in the most optimal way, using demographic information, to ensure a high SSR. The framework takes advantage of both quantitative and qualitative inputs (numerical and categorical), including:

- Surgery type (Total Hip Replacement, Total Knee Replacement, Hip Resurfacing, Unilateral knee replacement)
- Sex (The average surgery time will depend between male and female patients)(Farid Al Zoubi, Paul E. Beaulé, and Pascal Fallavollita, n.d.)
- Age of the patient
- BMI (BMI may not have a significant impact on surgery time but influences surgery success rates and readmissions)(Farid Al Zoubi, Paul E. Beaulé, and Pascal Fallavollita, n.d.)
- ASA patient classification (1 to 4) and anesthesia type. It serves as indicators of a patient's health status. ASA 1 signifies excellent health with no underlying illnesses, while ASA 4 indicates poor health with multiple chronic conditions.

The patient scheduling is completely independent of the operation itself and the medical professionals involved in the surgery, and focuses on arranging surgeries in the most optimal way possible. With proper model training the PSF can help a healthcare institution achieve two different types of positive outcomes. The first would be to group patients optimally. For example, if an institution has to divide two patients over five days of surgery with four patients each day, it can distribute patients that require more time (e.g.

males with high BMI for a more time-consuming procedure) over the available days instead of grouping them all together.

The second positive outcome would be to group the most time-consuming patients in one day. Sacrificing one day may help the surgical team achieve a better weekly success rate. It can be extrapolated for longer time frames (months, years, etc.), taking the fatigue of the surgical team into account. This framework is both easy to create and implement as the inputs are readily available. The upcoming section titled "Explanation of Models/Frameworks Output" elaborates on the outputs generated by this framework, providing illustrative examples for clarification.

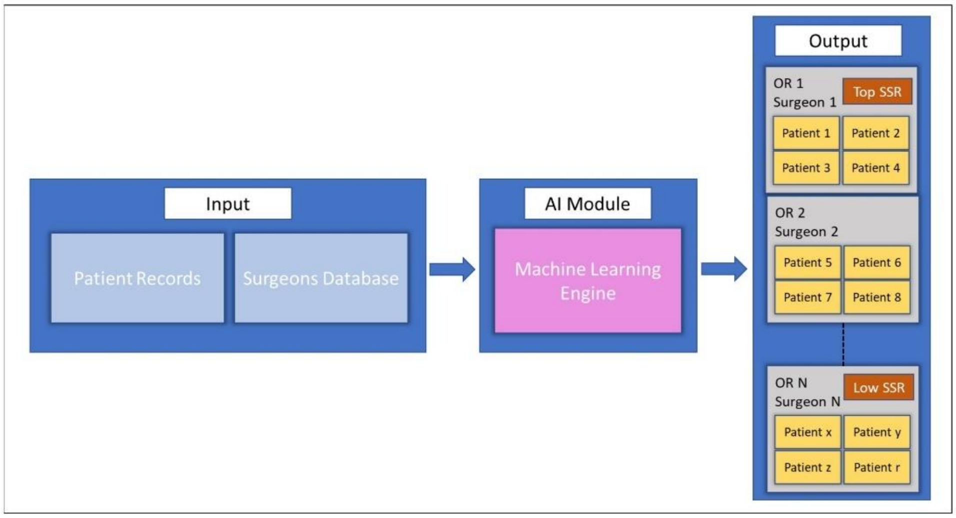


Figure 2: Patient Scheduling Framework. The ML Engine suggests patient groupings, with an emphasis on demographics, ranging from the highest to the lowest probability of success rate (SSR).

5.1.2 Surgical Team Scheduling Framework

The Surgical Team Scheduling Framework (STSF), Figure 3, focuses on identifying and utilizing the best surgical personnel combination to predict and influence the success rate of surgery. Surgical team scheduling for clinicians was developed using supervised machine learning models, which is atypical since segmentation and grouping are typically achieved through unsupervised learning methods, particularly clustering. For this paper, the success rate was defined as completing four arthroplasty surgeries within eight hours, but it can be extrapolated for almost all surgical procedures, including the ones that may be performed without general anesthesia. Inputs of the STSF are qualitative in nature i.e., categorical data. The three inputs are:

- Surgeons
- Anaesthesiologist
- Circulating nurses (i.e., registered nurses (RN))

Since the number of surgeons, RNs, and anaesthesiologists is usually pre-defined, the STSF requires minimal data and input. There is a limited pool of available variables to choose from, and it becomes even more constrained with explicit conditions like specific surgeons working on specific days or pre-defined assignments for RNs. The framework will strive to predict the highest degree of success by rearranging the available variables/personnel into different groups. The framework is naturally the most efficient because by combining the most efficient and compatible individuals in one group, you can naturally achieve better efficiency and productivity while forestalling problems that result in delays.

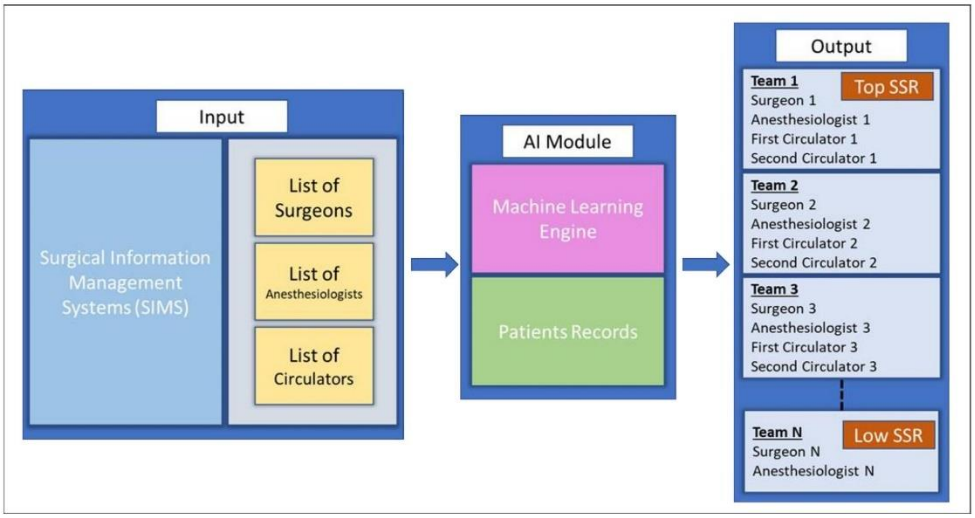


Figure 3: Surgical Team Scheduling Framework. The AI module reconciles inputs with patient-related factors and surgical team members then runs the ML engine to suggest the best team combinations.

5.1.3 Workflow/Time Monitoring Framework

The Workflow /Time Monitoring Framework (WTMF), Figure 4, aims to improve efficiency and help achieve a higher success rate by monitoring and optimizing the time taken by each step of a well-defined healthcare process/procedure. The arthroplasty procedures we compiled data on had the following workflow steps as shown in Figure 5:

1. Surgical Preparation Time (SPT)
2. Anesthesia ready
3. Surgical procedure duration (Procedure start time stamp to Procedure finish time stamp)
4. Anesthesia finish time (AFT)
5. Procedure finish to patient out of room time (OR specific time variables like cleanup, communication, post-surgery reflections, etc.)

6. Overall turnover time (It's governed by both internal and external factors, like delays in preparing the next patient for surgery that takes place outside the OR).

The output would be the time-benchmarks mapped to the probability of the successful completion of four surgeries in the given time for each set of benchmarks as explained in the "*Models/Frameworks Output Explanation*" section of the paper.

These six time-specific numerical inputs are logged by a circulating nurse in the SIMS of the hospital, which pre-emptively removes part of the bias associated with obtaining these variables (i.e. objective observer). These inputs are also what makes this framework procedure agnostic, as virtually every surgical procedure can be divided into multiple time-bound steps to identify the outliers.

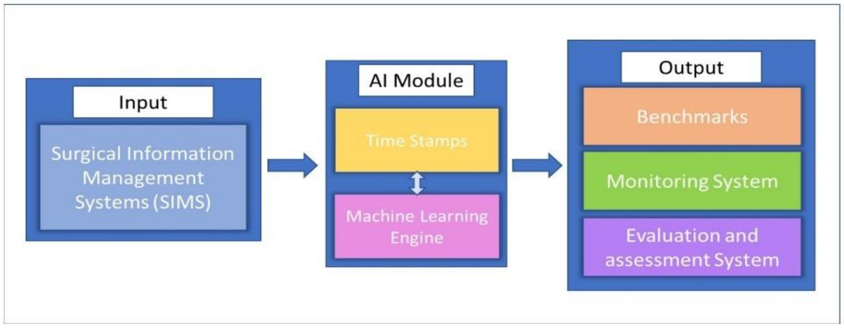


Figure 4: Workflow / Time Monitoring Framework. The ML engine/model leverages the timestamps to assist in the development of an efficient monitoring system that relies on efficiency/optimization benchmarks and an evaluation and assessment system.

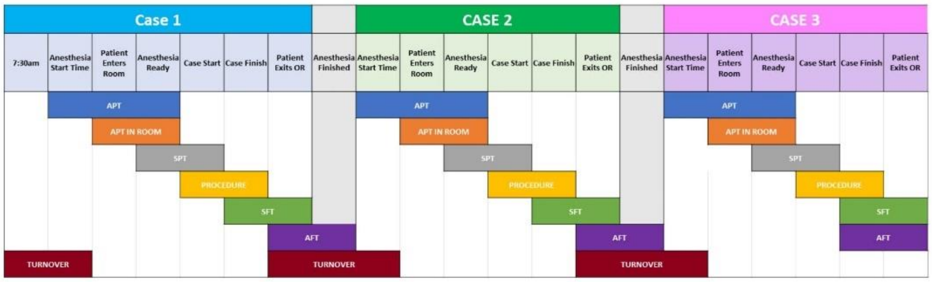


Figure 5: Surgical workflow steps for the first three surgical cases of a four joint surgery day. The fourth surgical case would have to end by 3:30pm for the day to be considered a success.

5.2 The ML-Models

We conduct a comparative analysis of the performance of supervised machine learning models employed in each framework. The evaluation metrics used for comparison encompass 6-fold Cross Validation (CV)

Accuracy, AUC-ROC, Sensitivity (Recall), Specificity, Precision, and we also assess the presence of Overfitting.

5.2.1 Why These Models.

We chose six machine learning algorithms to compare data collected for each of the three frameworks. The algorithms are:

- LR (Logistic Regression) (Nusinovici et al. 2020)
- SVM (Support Vector Machine) (Noble 2006)
- RF (Random Forest) (Biau and Scornet 2016)
- DNN-ANN (Deep Neural Network-Artificial Neural Networks with multiple hidden layer) (Montavon, Samek, and Müller 2018)
- XGBoost (Extreme Gradient Boosting: Gradient-Boosted Decision Tree (GBDT)) (Chen and He 2014)
- DT (Decision Tree) (Huang and Yeh 2006)

The rationale behind choosing six different models was to conduct a comprehensive ML analysis from various perspectives to identify the best framework and model outputs.

Parametric vs. Non-Parametric Models: We chose both parametric (like LR) and non-parametric models like (SVM and DT) to see how they would perform on our dataset. Even though the data is well-defined and categorized (better-suited for parametric models), the less-restrictive and flexible non-parametric models may identify patterns and connections eluded by parametric models (Clarke, Duda, and Hart 1974).

Ensemble Learning: Using six different models and covering ensembling techniques like bagging (via Random Forest), boosting (via XGBoost), and to an extent, stacking helped us achieve desired ensemble learning outcome, i.e., improving average performance prediction.

Interpretable Models: We also chose models like Decision Trees and Logistic Regression with a high degree of human comprehension. These allowed us to identify the most influencing metrics and data sources. Interoperability is also easier to refine and relatively easily understood by domain experts and not just data scientists. The WTMF can identify time-benchmarks as baseline for every stage of the surgery. When PSF predicts an unsuccessful day due to specific patients in a specific OR, surgeon don't have to reschedule all four surgeries but simply reroute a patient or patients that do not meet a specific demograhic criteria (Clarke, Duda, and Hart 1974).

Deep Learning: In order to cover all bases, we also used a Deep Neural Network (DNN-ANN) on our structured data.

5.2.2 Why These Evaluation Metrics

We chose six metrics to cover different ML performance evaluation aspects.

- **AUC-ROC**: is our measure of a model's performance from a classification perspective, and it evaluates performance for all possible classification thresholds.
- **Cross Validation (CV)**: We ran multiple cross-validation iterations (6 folds) on the data subsets to generate valid CV accuracy numbers, and it serves as the primary measure of the overall performance of a model
- **Sensitivity** and **specificity**, for per-class performance. Sensitivity or recall is how well (how frequently) a model recognizes true positives "successful days" out of total instances. While

specificity is a ML model's ability to identify a true negative "unsuccessful day" and is often used in conjunction with sensitivity to evaluate how accurate a model is.

- **Precision** shows the classifier's performance for class imbalanced data. Precision allows us to evaluate the quality of a ML model's positive predictions, successful days. From an (OR) perspective, this means that when a model predicts a day to be successful, precision is assessing how frequently the model's predictions align with the actual successful days.
- **Overfitting**: We measured the difference between training and testing accuracies as an indicator of the future error when the model undergoes a new dataset. If the model is overfitted, it may have trouble generalizing and adapting to the new data, which may result in inaccurate classifications and predictions.

6 Model Comparison and Results

This section is structured into six segments. In the initial three segments, we conduct individual model comparisons for each framework. The fourth segment comprises general observations and a comparison of models across all frameworks. The fifth segment is a discussion of the AI frameworks from the viewpoint of the models themselves. Lastly, we delve into explaining the outputs generated by the models and the frameworks. We provide examples of the expected outputs for each framework and offer guidance on how to interpret them based on the available data. Additionally, we illustrate how altering a specific metric value can influence the SSR, in other words, the OR throughput.

The parameters utilized for tuning the ML models can be found in Table 4 along with their corresponding values listed in the Appendix. Additionally, Table 8 presents the outcomes of each individual ML model and provides results when comparing models against one another.

Model	Parameters
LR	penalty, class_weight, solver
SVM	C, kernel, gamma, degree
RF	n_estimators, max_depth, min_samples_split, min_samples_leaf
DNN-ANN	hidden_units, estimator,max_epochs
XGBoost	loss, n_estimators, criterion
DT	criterion, splitter, max_depth, min_samples_split, min_samples_leaf

Table 4: Parameters used to tune the ML Models. The values for each parameter were obtained using a Grid Search process

6.1 Patient Scheduling Models

For patient scheduling, the CV accuracy ranged between 67% (highest) and 57% (lowest), with a standard deviation of 0.04 (4%). LR had the highest accuracy, while SVM ranked lowest. This means we can predict successful days 67% ± 4% of the time.

DT was identified as the ideal ML model for patient scheduling for the following reasons:

- Second highest CV accuracy score, just one percent below the best score
- It's one of the two least over-fitted models for this solution. The other one was DNN-ANN, but it had a significantly lower precision, CV accuracy, and AUC-ROC score.
- It's one of the two most precise models. The other one was RF, but it was more overfitted and less accurate than DT.
- Highest specificity.

Despite the highest CV accuracy score, LR wasn't chosen as the best ML model fit for this solution because of a high overfitting and lower precision score. The chosen model's high sensitivity and high specificity scores indicate that it will offer very few false positives.

6.2 Surgical Team Scheduling Models

The CV accuracy range for the six ML models is from 59% (lowest) to 85% (highest), and the standard deviation of 0.1 (10%) is significantly higher compared to the previous framework. For surgical team scheduling, the DT model had the highest CV accuracy, while LR scored the lowest. With 85% success rate, we can confidently predict days that are going to be successful accurately.

However, the ideal ML model fit for this solution was the one with the second highest CV accuracy score, i.e., XGBoost. The other reasons why it was identified as the best fit are as follows:

- Third-least overfitted model and far closer to the lowest range than the highest one.
- Precision is close to the most precise ML model (DT).
- Sensitivity among the top 3 models and average specificity.
- Second-highest AUC-ROC score.

The model with the highest accuracy (DT) was one of the two most poorly fitted ones though still not as overfitted as DNN-ANN. XGBoost is a well-rounded fit. It's quite precise, very accurate compared, and modestly sensitive. For this solution, the choice was really among four models, i.e., DT, XGBoost, RF, and SVM, as the other two negative outliers with poor scores for almost all metrics.

6.3 Workflow /Time Monitoring Models

For workflow / time monitoring, the CV Accuracy, which we have taken as the primary measure of the performance of an ML model, ranged between 68% (lowest) and 76% (highest). The standard deviation of 0.03 (3%) is relatively minimal. This implies that 76% ± 3% of the time we can predict successful days given the Time-Metrics as input for our model.

However, the ideal ML model for Time Monitoring data is not LR, with the highest CV accuracy level. The best ML model for Time Monitoring is Random Forest for the following reasons:

- Least overfitted model.
- CV accuracy is exactly in the middle of the range.
- Highest precision.

LR had the highest CV accuracy score, but it was also the most overfitted model out of the six examined.

SVM was second in line from the CV accuracy perspective, but it was the least precise of all models. Thus, LR proved to be the most well-rounded pick as the top ML model for this solution, even though its sensitivity and specificity scores were among the lowest. This indicates that the model might not flag as many true positives and negatives as other models, but the ones it picks are likely to be accurate four out of five times. Thanks to low overfitting, it's also expected to be more flexible and adaptable to new data.

6.4 Additional observations

An interesting observation regarding the data type for each solution and its impact on the CV accuracy of the model was that Logistic Regression is *most* accurate with numerical data and *least* accurate with categorical data inputs. On the other end of the spectrum, Decision Trees are *least* accurate with numerical data and most accurate with categorical data. The fact that its CV accuracy score is highest for mixed data (numerical and categorical) and DT's accuracy is second highest for mixed data endorses this conclusion. SVM has the lowest accuracy score for mixed data. There are several other patterns and trends that were observed when we evaluated different ML models for each framework. Some of the most significant ones are highlighted below.

Overfitting

If we evaluate different models from an overfitting perspective, the following trends emerge:

- **LR** is most likely to overfit with purely numerical data sets. The degree of overfitting is minimal with purely categorical data sets and averages for data sets that include both numerical and categorical inputs.
- **DT** is prone to a high degree of overfitting when data is either purely numerical or categorical. However, it doesn't overfit at all when the data is mixed.
- **XGBoost** is most likely to overfit with mixed type of data. The probability of overfitting with numerical data might be 50/50, while it's least likely to overfit with categorical data. The overfitting is also directly proportional to model complexity.
- **RF** is more likely to overfit with purely categorical data compared to purely numerical and mixed data.
- **SVM** is more likely to overfit when there is not enough data, or there is a high degree of misclassification. From our observation, it was more likely to be numerical data than categorical data.
- **Overfitting** in **DNN-ANN** was uncharacteristically high when dealing with purely categorical data and least likely to overfit with mixed data. Its overfitting tendencies can be connected to model complexity but showed the opposite relationship to XGBoost when it comes to data types and model complexity.

As for the solutions, it was observed that ML models trained on patient scheduling data (mixed between numerical and categorical) are least likely to overfit. In contrast, models trained on surgical team scheduling are most likely to overfit and to a much higher degree in comparison. The probability of overfitting for models trained in workflow/time monitoring data is closer to models trained on patient scheduling and showed a similar standard deviation.

Precision: A higher degree of precision is sought after in ML models because it's more important to have accurate predictions (successful days in our case) than more predictions (both successful and unsuccessful). Accurate predictions augment the reliability of an ML model and can be used for decision-making with more certainty.

DNN-ANN: Despite the level of complexity and sophistication it offers, the DNN-ANN model adds no value to our comparison and has relatively low scores for all metrics in all three solutions.

6.5 AI-Frameworks from the ML models' perspective

Patient Scheduling: It's the least accurate model, but it's also the most flexible to new data as ML models trained in patient scheduling data are least likely to overfit. Its sensitivity is low across the board. However, its precision numbers are similar for all ML models.

Surgical Team Scheduling: This framework shows bias against two ML models - LR and DNN-ANN, for which the numbers are disproportionately poor. But even if we identify trends in the remaining four models, surgical team scheduling is most prone to overfitting, especially if you choose the wrong model. However, it also offers the most accurate results. It also has the best AUC-ROC, Sensitivity, Specificity, and Precision numbers, if you discount DNN-ANN and LR from consideration.

Workflow/Time Monitoring: Medium accuracy and a relatively low chance of retrofitting compared to the other two frameworks. Considering its low CV accuracy standard deviation, it's the most robust of the three models. Classification problems and sensitivity might be similar across all ML models trained on the numerical time modeling data.

6.6 FAST Output Explanation

Only a limited number of supervised machine learning models offer explainability (Belle and Papantonis 2021). We have opted for the decision tree model to illuminate the outputs of the models within each framework below. It is crucial to emphasize that the decision tree (DT) may not necessarily be the best-performing model for each framework. Our choice of the DT model is solely for the purpose of illustration and to enhance comprehension of the inter-relationships among various elements. It's worth noting that, for the sake of clarity, we have not included all potential metrics for each model in our examples below. Instead, we have chosen specific metrics. Otherwise, the number of scenarios and model output combinations is nearly limitless, leading to various possible SSR outcomes.

Before we examine the scenarios ahead, it's crucial to make a clear distinction between the outputs generated by the models and those produced by the framework. The model's output comprises predictions regarding the likelihood of a day being successful (i.e., completing four surgeries before 3:30 pm). On the other hand, the framework's output consists of patient groups, surgical teams, or workflow/time benchmarks necessary to attain specific SSR goals during a surgical day.

Patient Scheduling Scenarios

The concept behind the patient scheduler is to arrange patients into groups of four based on their demographics in a manner that maximizes the Surgical Success Rate (SSR) to the greatest extent possible. Table 5 presents multiple scenarios depicting the impact on SSR when modifying a single patient metric at a time. For instance, Scenario 1 illustrates that scheduling four healthy women with ASA scores of 1 or 2, who are also young (aged less than 42), for hip arthroplasty procedures in a single day, increases the likelihood of completing the fourth surgery before 3:30 pm, thereby avoiding overtime pay, with a high probability of 91%. Conversely, when considering group patients with similar characteristics but in poorer health (ASA 3 and 4)., the SSR remains astonishingly high at 86%

Likewise, in scenarios 3 and 4, an increase in patient age resulted in only a marginal 3% reduction in SSR. The same trend holds true for anesthesia type in scenarios 5 and 6. What is noteworthy here is that a higher proportion of males in an operating room is associated with a lower SSR. When we compare scenarios 1 and 6, it becomes evident that by substituting four females with four males who share similar patient characteristics and undergo the same procedure, we are significantly compromising our OR throughput, with a nearly 40% decrease. This indicates to OR administrators that scheduling more than half of the patients as males in a single day may not be advisable. Similarly, the scheduler should have the capability to provide suggestions regarding the maximum allowable combined age or BMI sum for the patient group to be scheduled in a single day. In cases where such a grouping is necessary, it may be advisable to consider selecting all females or patients undergoing specific procedures. Authors in (Farid Al Zoubi, Paul E. Beaulé, and Pascal Fallavollita, n.d.) provides insights into the patient-related factors that contribute to delays in arthroplasty surgeries. These insights can aid in comprehending the reasoning behind certain scheduling strategies employed by this framework.

Scenario	Surgery	Anesthesia	Gender	Age	ASA	SSR
1	HR, THA	Spinal	4F	<42	1,2	91%
2	HR, THA	Spinal	4F	<42	3,4	86%
3	UKA	Spinal	3F, 1M	<57	Any	67%
4	UKA	Spinal	2F,2M	58-70	Any	64%
5	THA	Spinal	4M	42-53	1,2	58%
6	THA	General	4M	42-53	1,2	54%

Table 5: SSR for different patient scheduling scenarios

Surgical Team Scheduling Scenarios

Surgical team scheduling primarily centers around organizing the operational team, which comprises circulators, surgeons, and anesthesiologists. Scenario 4 presented in Table 6 below illustrates that there is an inevitability of incurring overtime costs on days when a particular team, consisting of Surgeon D, Circulator CC, and Anesthesiologist V, is assigned to work together. However, the model indicates that if Anesthesiologist V is substituted with Anesthesiologist L, the SSR increases significantly to 92%. In contrast, scenarios 1 and 2 do not exhibit the same pattern, as replacing the anesthesiologist in those cases leads to only a 5% alteration in the SSR. Indeed, this demonstrates the significant impact that a single team member can have on the overall efficiency of the team. The team scheduler's role is to analyze the team's data and provide recommendations for creating the most productive group to operative on a specific patient.

Scenario	Surgeon	Circulator 1	Anesthesiologist	SSR
1	D	CC	L	92%
2	D	CC	P	87%
3	D	GG	V	65%
4	D	CC	V	0%

Table 6: SSR for different team scheduling scenarios

Workflow /Time Monitoring Scenarios

The output from the workflow/time monitoring framework exhibits some differences from the preceding two frameworks. In this framework, data input is expected to flow in a cascading manner, with the output of one stage influencing the subsequent one (Figure 5). The machine learning model must continually adapt the SSR during the course of the operational time, operating under the assumption that forthcoming stages have predefined baseline values (Al Zoubi et al. 2022). This is why it's referred to as monitoring, as the surgeon or the OR administrator is expected to closely observe the time and react accordingly. Table 7 presented below demonstrates that when the turnover process between two consecutive surgeries exceeds 21.5 minutes, it can lead to a significant decrease in the SSR, from 69% to 59%, as in scenarios 3 and 4. However, this is not the situation when a comparable delay of 11 minutes occurs within the surgery itself, during the surgical procedure, scenarios 1 and 2. In simple terms, a one-minute delay in one stage of the surgery can have a vastly different impact on the SSR than in another stage. This complexity makes it challenging for a human to assess the significance of a minute at a specific point in the surgery without relying on the framework. As an additional motivation of implementing this framework, in our previous work (Farid Al Zoubi 2023) described the method of utilizing time saved from similar surgeries to accommodate additional procedures in a day, thereby reducing the arthroplasty waitlist.

Scenario	APT	Case	AFT	Turnover	SSR
1	<10.5	53	<20.5	<21.5	93%
2	<10.5	64	<20.5	<21.5	89%
3	<10.5	<71.5	<20.5	<21.5	59%
4	<10.5	<71.5	<20.5	>21.5	69%
5	10.5-18.5	<71.5	<20.5	<21.5	64%

Table 7: SSR for different stages-benchmarks Time monitoring scenarios

Patient Scheduling (Numerical and Categorical)						
Model	CV Accuracy	AUC-ROC	Sensitivity	Specificity	Precision	Overfitting
LR	**67%**	63%	19%	86%	57%	3%
DT	66%	65%	50%	97%	64%	1%
XGBoost	65%	92%	37%	71%	53%	5%
RF	62%	58%	36%	72%	64%	3%
DNN-ANN	62%	58%	17%	86%	56%	1%

SVM	57%	53%	9%	88%	60%	4%
Standard Deviation	0.04	0.14	0.14	0.10	0.04	0.02

Surgical team Scheduling (Categorical)						
Model	CV Accuracy	AUC-ROC	Sensitivity	Specificity	Precision	Overfitting
DT	85%	93%	86%	80%	82%	11%
XGBoost	80%	92%	83%	79%	79%	3%
RF	74%	92%	85%	79%	81%	6%
SVM	71%	88%	82%	81%	80%	1%
DNN-ANN	62%	50%	49%	68%	59%	20%
LR	59%	54%	0%	100%	31%	2%
Standard Deviation	0.10	0.20	0.34	0.10	0.20	0.07
Workflow /Time Monitoring (Numerical)						
Model	CV Accuracy	AUC-ROC	Sensitivity (Recall)	Specificity	Precision	Overfitting
LR	76%	74%	74%	84%	62%	8%
SVM	75%	81%	74%	77%	27%	6%
RF	72%	72%	66%	59%	80%	1%
DNN-ANN	70%	74%	73%	59%	64%	4%
XGBoost	70%	76%	72%	61%	60%	4%
DT	68%	67%	74%	55%	53%	5%
Standard Deviation	0.03	0.04	0.03	0.12	0.17	0.02

Table 8: Models and their performance compared. The best ML model for each framework has its row highlighted in green

7 Discussion

Instead of adding resources (human, financial, technical, etc.) to improve OR efficiency and complete more surgeries on time, we proposed FAST; based on data-driven solutions to efficiently use the available resources. This required us to leverage data, artificial intelligence, and multiple machine learning models to modify/improve our processes and healthcare delivery practices to achieve a better success rate. Quality of patient care and patient safety remained our top priority throughout our pursuit of OR optimization solutions that could be repurposed for other surgical procedures and other elements of healthcare delivery. The goal was to develop good practices that can be adopted by a diverse range of surgical teams and healthcare facilities.

This work goes beyond the predictive analytics, even though the frameworks we have developed focus on predictions. FAST are precursors to comprehensive prescriptive frameworks, which can evolve with the right additions and development. Unlike frameworks based on only on predictive analytics, prescriptive frameworks don't just identify past patterns. Instead, they leverage past data to make actionable predictions that can be implemented to better control or improve future outcomes. To our knowledge, this work presents a first when conceiving AI-driven frameworks to improve efficiency and productivity that take into consideration the following:

1. They entail the collaboration of all key stakeholders, including patients' demographics, surgical professionals, and time management.
2. Encompass all stages of the surgical process, including pre-operation, intra-operation, and post-operation phases
3. They are stand-alone solutions that can be implemented individually or in various combinations.
4. They leverage ML models with diverse data input streams, making the complete set suitable for a wide range of healthcare institutions. This allows institutions to choose the framework that aligns best with their available data.
5. The AI-driven frameworks proposed in this study were developed and validated using orthopedic patient-specific data.
6. These adaptable frameworks can also be applied to other areas of application.

We identified WTMF to be the farmwork with the most potential to significantly impact the surgical success rate (SSR) of different surgeries because optimizing even a single step in a surgical procedure for efficiency can directly impact the result, i.e., completion time. Similarly, even small adjustments in individual steps can lead to substantial collective time reduction, leading to a high success rate, and a shorter arthroplasty waitlist (Farid Al Zoubi 2023). Using this framework and integrating simple recommendations from positive outliers, i.e., most time-efficient professionals, our clinical institution managed to improve the arthroplasty surgical success rate from 39% to 72% (output) (Al Zoubi et al. 2022). However, the opposite is just as true. A significant delay, whether caused by one team member or multiple individuals, can prevent the timely completion of healthcare procedures within the given time frame. It requires the whole team to work efficiently, which, in turn, requires better communication protocols, compatibility, aligning work styles, etc. The cost of building and applying such a framework can be significantly high. In conclusion, despite its complexity, cost, delicacy, and implementation requirements, the WTMF may significantly and immediately impact overall team efficiency. In our previous work (Al Zoubi et al. 2022), we have worked on improving this framework and transforming it into a prescriptive ML framework utilizing the Positive Deviance technique.

Once we have accumulated statistically significant data, we can expand out of just one framework and combine multiple solutions to develop one comprehensive perspective framework that can incorporate the strengths of all individual solutions. Such a framework may help surgical teams schedule patients, identify optimal teams/healthcare professional combinations, and offer time monitoring/optimization insights to significantly increase the probability of all surgeries being completed on time for any given day. This will ultimately lead to a higher OR throughput.

8 Challenges and Limitations

One of the most significant challenges in implementing these solutions is their multidisciplinary requirements. This starts with the healthcare itself, where overlap among different professionals (surgeons, nurses, and anesthesiologists) is relatively minimal. While professionals from the three different healthcare disciplines routinely interact with each other during procedures, there are inherent professional boundaries that may prevent a seamless adoption of a solution involving all three. For example, some nursing unions won't allow their nurses to begin work prior to &:30am in the morning, which would inherently add preparation time to an already busy surgical day. Then there is the engineering layer connected to the development and deployment of these frameworks. Engineers may perceive the inefficiencies and possible solutions from an entirely different perspective than healthcare professionals. Business owners have unique considerations and reservations compared to engineers and healthcare professionals, ranging from financial to human resource management. The conflicts of interest of these three healthcare stakeholders may slow down or impede the adoption and deployment of these frameworks in a healthcare institution. Getting them on the same page regarding one or multiple framework(s) for OR optimization will be a major challenge.

A major limitation inherent to this body of work is the inherent margin of error associated with uncertainty and ambiguity. Both epistemic and aleatoric uncertainty (dimensions) is worth considering. Epistemic uncertainty is common in ML models with limited, incomplete, or inappropriate training data. This uncertainty can be partially eliminated by improving the training data available. In contrast, aleatoric uncertainty is tied to measurement errors and randomness that can't be explained away. You can quantify and account for it, but it's difficult to reduce, let alone eliminate, even with the introduction of more data. Then there is the inherent ambiguity of the frameworks. There is always a possibility that a different framework or the existing framework with certain variations may lead to a better interpretation of available data to produce more potent, actionable, efficient, or less resource-intensive solutions. Nonetheless, we believe that the research community can expand on FAST to best suit their needs.

Some frameworks also have certain drawbacks. The surgical team scheduling framework, for instance, can lead to team polarization. It will create two extremes, i.e., the strongest and the weakest members of the staff, which may lead to further division. The best groups will be becoming stronger together, while the weaker ones may not get adequate chances to learn from the best, which they would have gotten from random or deliberate team assignments. It may also lead to team resistance when surgeons and their preferred teams are disassembled and reassembled. The patient scheduling framework has drawbacks of its own, starting with the assigned surgeon. If the surgeon that must perform a surgery is not available for the day when the patient should optimally be scheduled (per the framework), it's a constraint. Working around this constraint may significantly reduce the efficiency of the framework. Moreover, the framework showed the lowest success rate amongst other frameworks indicating that patient metrics are not crucial to the success rate when it is compared with the operating staff.

9 Conclusion

When it comes to making arthroplasty surgeries more efficient, AI-driven frameworks, using the right ML models, can reconcile a wide range of time, individual (surgeons, patients, etc.), and scheduling variables to identify patterns and test them against a diverse range of benchmarks to predict and improve the surgical success rate SSR. The insights they generate can be applied to different phases of the surgical procedures (from pre-op to post-op), which collectively increase the possibility of on-time completions and higher operating rooms throughput at lower cost.

Different ML algorithms were applied to the available arthroplasty dataset, and their performance was evaluated based on metrics like accuracy, sensitivity, and overfitting. However, the choice for the right ML model was governed by far more than just their performance numbers. Multiple factors, from the difficulty of obtaining relevant inputs to the potency of the optimization insights they helped generate, were considered. This rational choice of ML models was one of the core contributors to the successful development of three comprehensive decision-support AI-frameworks capable of improving surgical workflow processes. FAST can react to real-time changes in variables and offering insights, which, if applied to the subsequent surgical stages, may still lead to the successful completion of surgeries on time. Even though the frameworks were developed using data obtained from our clinical institution, they can be leveraged to improve OR efficiency and increase OR throughput for a variety of surgical procedures.

10 References

Bartek, Matthew A., Rajeev C. Saxena, Stuart Solomon, Christine T. Fong, Lakshmana D. Behara, Ravitheja Venigandla, Kalyani Velagapudi, John D. Lang, and Bala G. Nair. 2019. "Improving Operating Room Efficiency: Machine Learning Approach to Predict Case-Time Duration." *Journal of the American College of Surgeons* 229 (4): 346-354e3. https://doi.org/10.1016/j.jamcollsurg.2019.05.029.

Beaulé, Paule E., Aaron A. Frombach, and Jae-Jin Ryu. 2015. "Working toward Benchmarks in Orthopedic OR Efficiency for Joint Replacement Surgery in an Academic Centre." *Canadian Journal of Surgery* 58 (6): 408–13. https://doi.org/10.1503/cjs.001215.

Belle, Vaishak, and Ioannis Papantonis. 2021. "Principles and Practice of Explainable Machine Learning." *Frontiers in Big Data*. https://doi.org/10.3389/fdata.2021.688969.

Biau, Gérard, and Erwan Scornet. 2016. "A Random Forest Guided Tour." *Test* 25 (2). https://doi.org/10.1007/s11749-016-0481-7.

C. Yeoh, J. Mascarenhas, K.S. Tan, and L. Tollinche. 2018. "Real-Time Locating Systems and the Effects on Efficiency of Anesthesiologists ." *Journal of Clinical Anesthesia and Pain Management* 2, no. 1.

"Canadian Institute for Health Information. (2020) Hip and Knee Replacements in Canada: CJRR Quick Stats, 2018–2019. Ottawa, ON." n.d.

Chen, Tianqi, and Tong He. 2014. "Xgboost: Extreme Gradient Boosting." *R Lecture*, no. 2016.

Clarke, M. R. B., Richard O. Duda, and Peter E. Hart. 1974. "Pattern Classification and Scene Analysis." *Journal of the Royal Statistical Society. Series A (General)* 137 (3). https://doi.org/10.2307/2344977.

Cram, Peter, Bruce E. Landon, John Matelski, Vicki Ling, Therese A. Stukel, J. Michael Paterson, Rajiv Gandhi, Gillian A. Hawker, and Bheeshma Ravi. 2018. "Utilization and Short-Term Outcomes of Primary Total Hip and Knee Arthroplasty in the United States and Canada." *Arthritis & Rheumatology* 70 (4): 547–54. https://doi.org/10.1002/art.40407.

Dehnoei, Sajjad, Antoine Sauré, Onur Ozturk, William Gardner, Kathleen Pajer, Roxanna Sheppard, and Jonathan Patrick. 2022. “A Stochastic Optimization Approach for Staff Scheduling Decisions at Inpatient Units.” *International Transactions in Operational Research*. https://doi.org/10.1111/itor.13226.

Erekat, Asala, Gregory Servis, Sreenath Chalil Madathil, and Mohammad T. Khasawneh. 2020. “Efficient Operating Room Planning Using an Ensemble Learning Approach to Predict Surgery Cancellations.” *IISE Transactions on Healthcare Systems Engineering* 10 (1): 18–32. https://doi.org/10.1080/24725579.2019.1641576.

Eshghali, Masoud, Devika Kannan, Navid Salmanzadeh-Meydani, and Amir Mohammad Esmaieeli Sikaroudi. 2023. “Machine Learning Based Integrated Scheduling and Rescheduling for Elective and Emergency Patients in the Operating Theatre.” *Annals of Operations Research*. https://doi.org/10.1007/s10479-023-05168-x.

Fairley, Michael, David Scheinker, and Margaret L. Brandeau. 2019. “Improving the Efficiency of the Operating Room Environment with an Optimization and Machine Learning Model.” *Health Care Management Science* 22 (4): 756–67. https://doi.org/10.1007/s10729-018-9457-3.

Farid Al Zoubi, Pascal Fallavollita Georges Khalaf, Paul E. Beaulé. 2023. “Leveraging Machine Learning and Prescriptive Analytics to Improve Operating Room Throughput.” *Frontiers in Digital Health* 5.

Farid Al Zoubi, Paul E. Beaulé, and Pascal Fallavollita. n.d. “Factors Influencing Delays and Overtime During Surgery: A Descriptive Analytics for High Volume Arthroplasty Procedures.”

Garg, Swati, Shuchi Sinha, Arpan Kumar Kar, and Mauricio Mani. 2022. “A Review of Machine Learning Applications in Human Resource Management.” *International Journal of Productivity and Performance Management*. https://doi.org/10.1108/IJPPM-08-2020-0427.

Huang, Hsiu Li, and Mei Chang Yeh. 2006. “Introduction to Ethnographic Decision Tree Modeling.” *Journal of Nursing*.

Jain, Abhinav, Hima Patel, Lokesh Nagalapatti, Nitin Gupta, Sameep Mehta, Shanmukha Guttula, Shashank Mujumdar, Shazia Afzal, Ruhi Sharma Mittal, and Vitobha Munigala. 2020. “Overview and Importance of Data Quality for Machine Learning Tasks.” In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. https://doi.org/10.1145/3394486.3406477.

Jordan, M. I., and T. M. Mitchell. 2015. “Machine Learning: Trends, Perspectives, and Prospects.” *Science*. https://doi.org/10.1126/science.aaa8415.

Lazebnik, Teddy. 2023. “Data-Driven Hospitals Staff and Resources Allocation Using Agent-Based Simulation and Deep Reinforcement Learning.” *Engineering Applications of Artificial Intelligence* 126. https://doi.org/10.1016/j.engappai.2023.106783.

Maheshwari, Kamal, Jing You, Kenneth C. Cummings, Maged Argalious, Daniel I. Sessler, Andrea Kurz, and Jacek Cywinski. 2017. “Attempted Development of a Tool to Predict Anesthesia

Preparation Time From Patient-Related and Procedure-Related Characteristics." *Anesthesia & Analgesia* 125 (2): 580–92. https://doi.org/10.1213/ANE.0000000000002018.

Maina, Elizaphan M., Robert O. Oboko, and Peter W. Waiganjo. 2017. "Using Machine Learning Techniques to Support Group Formation in an Online Collaborative Learning Environment." *International Journal of Intelligent Systems and Applications* 9 (3). https://doi.org/10.5815/ijisa.2017.03.04.

Montavon, Grégoire, Wojciech Samek, and Klaus Robert Müller. 2018. "Methods for Interpreting and Understanding Deep Neural Networks." *Digital Signal Processing: A Review Journal*. https://doi.org/10.1016/j.dsp.2017.10.011.

Myszczynska, Monika A., Poojitha N. Ojamies, Alix M.B. Lacoste, Daniel Neil, Amir Saffari, Richard Mead, Guillaume M. Hautbergue, Joanna D. Holbrook, and Laura Ferraiuolo. 2020. "Applications of Machine Learning to Diagnosis and Treatment of Neurodegenerative Diseases." *Nature Reviews Neurology*. https://doi.org/10.1038/s41582-020-0377-8.

Noble, William S. 2006. "What Is a Support Vector Machine?" *Nature Biotechnology*. https://doi.org/10.1038/nbt1206-1565.

Nusinovici, Simon, Yih Chung Tham, Marco Yu Chak Yan, Daniel Shu Wei Ting, Jialiang Li, Charumathi Sabanayagam, Tien Yin Wong, and Ching Yu Cheng. 2020. "Logistic Regression Was as Good as Machine Learning for Predicting Major Chronic Diseases." *Journal of Clinical Epidemiology* 122. https://doi.org/10.1016/j.jclinepi.2020.03.002.

Porta, Christopher R., Andrew Foster, Marlin W. Causey, Patricia Cordier, Roger Ozbirn, Stephen Bolt, Dennis Allison, and Robert Rush. 2013. "Operating Room Efficiency Improvement after Implementation of a Postoperative Team Assessment." *Journal of Surgical Research* 180 (1): 15–20. https://doi.org/10.1016/j.jss.2012.12.004.

Richard Gold, Farid Al Zoubi, Julia Brillinger, Kreviazuk Cheryl, Dennis Garvin, David Schramm, Pascal Fallavollita, Andrew Seely, Paul Beaulé. 2022. "Use of Multidisciplinary Positive Deviance Seminars to Improve Efficiency in a High-Volume Arthroplasty Practice; a Pilot Study." *Canadian Journal of Surgery*.

Rudra Kumar, M., Rashmi Pathak, and Vinit Kumar Gunjan. 2022. "Machine Learning-Based Project Resource Allocation Fitment Analysis System (ML-PRAFS)." In *Lecture Notes in Electrical Engineering*. Vol. 834. https://doi.org/10.1007/978-981-16-8484-5_1.

Salah, Haya, and Sharan Srinivas. 2022. "Predict, Then Schedule: Prescriptive Analytics Approach for Machine Learning-Enabled Sequential Clinical Scheduling." *Computers and Industrial Engineering* 169. https://doi.org/10.1016/j.cie.2022.108270.

Samorani, Michele, and Linda Goler Blount. 2020. "Machine Learning and Medical Appointment Scheduling: Creating and Perpetuating Inequalities in Access to Health Care." *American Journal of Public Health*. https://doi.org/10.2105/AJPH.2020.305570.

Schiele, Julian, Thomas Koperna, and Jens O. Brunner. 2021. “Predicting Intensive Care Unit Bed Occupancy for Integrated Operating Room Scheduling via Neural Networks.” *Naval Research Logistics (NRL)* 68 (1): 65–88. https://doi.org/10.1002/nav.21929.

Sculley, D., Gary Holt, Daniel Golovin, Eugene Davydov, Todd Phillips, Dietmar Ebner, Vinay Chaudhary, Michael Young, Jean François Crespo, and Dan Dennison. 2015. “Hidden Technical Debt in Machine Learning Systems.” In *Advances in Neural Information Processing Systems*. Vol. 2015-January.

Shohreh Majd; Zohreh Majd. 2023. “CLINICAL AND NON-CLINICAL MANAGEMENT IN HEALTHCARE: A COMPREHENSIVE OVERVIEW.” *International Journal of Advanced Research*.

Waly, Feras J., Donald S. Garbuz, Nelson V. Greidanus, Clive P. Duncan, and Bassam A. Masri. 2020. “Safety of a ‘Swing Room’ Surgery Model at a High-Volume Hip and Knee Arthroplasty Centre.” *The Bone & Joint Journal* 102-B (7_Supple_B): 112–15. https://doi.org/10.1302/0301-620X.102B7.BJJ-2019-1536.R1.

Zoubi, Farid Al, Richard Gold, Stéphane Poitras, Cheryl Kreviazuk, Julia Brillinger, Pascal Fallavollita, and Paul E. Beaulé. 2022. “Artificial Intelligence-Driven Prescriptive Model to Optimize Team Efficiency in a High-Volume Primary Arthroplasty Practice.” *International Orthopaedics*, June. https://doi.org/10.1007/s00264-022-05475-1.

11 Appendix

Framework	Model	Parameters			
PSF	DT	Criterion	splitter	max_depth	min_samples_split
		gini	best	3	2
	RF	n_estimators	max_depth	min_samples_split	criterion
		100	40	5	gini
	SVM	C	kernel	gamma	degree
		1	rbf	1	3
	LR	max_iter	C	solver	verbose
		100	5	liblinear	1
	DNN-ANN	hidden_layer_sizes	Activation	solver	
		1000	relu	adam	
	XGBoost	loss	n_estimators	criterion	
		log_loss	200	friedman_mse	
STSF	DT	Criterion	splitter	max_depth	min_samples_split
		entropy	best	16	5
	RF	n_estimators	max_depth	min_samples_split	criterion
		15	40	10	gini
	SVM	C	kernel	gamma	degree
		10	poly	0.1	3
	LR	max_iter	C	solver	verbose
		100	5	liblinear	1
	DNN-ANN	hidden_layer_sizes	Activation	solver	
		100	relu	adam	
	XGBoost	loss	n_estimators	criterion	
		log_loss	200	friedman_mse	
WTMF	DT	Criterion	splitter	max_depth	min_samples_split
		entropy	best	16	40
	RF	n_estimators	max_depth	min_samples_split	criterion
		100	40	5	gini
	SVM	C	kernel	gamma	degree
		1	rbf	1	3
	LR	max_iter	C	solver	verbose
		100	5	liblinear	1
	DNN-ANN	hidden_layer_sizes	Activation	solver	
		10	tanh	adam	
	XGBoost	loss	n_estimators	criterion	
		log_loss	50	friedman_mse	

Chapter 4: Use of multidisciplinary positive deviance seminars to improve efficiency in a high-volume arthroplasty practice: a pilot study

4.1. Summary

As previously mentioned, a crucial component within the general framework of any prescriptive analytics system involves identifying a specific set of actions to be examined on top of the output generated by predictive analytics, commonly known as "what-if" scenarios. This scholarly article focuses on gathering a diverse range of practical actions from various disciplines to streamline the process of surgeries, while ensuring the safety of patients remains uncompromised. Specifically targeting different groups of healthcare professionals such as nurses and surgeons, the study employs data analysis techniques to identify instances of positive deviance among all members, subsequently leveraging their best practices. After reaching a consensus on the most impactful actions, these recommendations are collected and can be seamlessly integrated into the prescriptive analytics systems.

4.2. Methodology

This chapter services as the connecting chain between the predictive analytics and the prescriptive analytics. This happens through introducing a recommendation system that works to bring the outcomes to certain value, in our case SSR. The primary objective of this thesis is to improve the operating room efficiency through improving team performance. In order to foster self-improvement and staff engagement to work as a team, various models of team efficiency have been developed using the LEAN method, Six-Sigma and process mapping which can be quite effective but very resource intensive. However, PD seminars use individual performance feedback to identify team members who outperform their peers. The strategies from those who demonstrate exemplary performance are used to both motivate peers and improve the practices for all. PD seminars focuses on individual strengths and resources already present, instead of focusing on negatives which require improvement. Implementing the strategies is feasible and sustainable as they are already in place and successful.

To assess the effectiveness of positive deviance sessions, a prospective cohort study was conducted, testing strategies as hypotheses to enhance performance, motivation, and teamwork in the operating room. Time interval data, patient demographics, adverse events, and other relevant details were recorded using the Surgical Information Management Systems (SIMS). The study focused on operating room efficiency using defined time intervals and involved five high-volume primary arthroplasty surgeons and 29 participating nurses. Positive deviants, identified through structured group interviews, led sessions to discuss strategies for improving time efficiency. These measures were implemented based on recommendations from combined sessions with nursing and surgeons, moderated by an experienced non-orthopedic surgeon. The study aimed to compare operating room efficiency before and after interventions to determine their success.

Two distinct sets of seminars were organized, with one set specifically designed for surgeons and another set tailored for nurses. Recommendations were compiled separately for the surgeons and nurses in their respective seminar sessions. Nevertheless, the assessment of the recommendations' implementation involved treating the before and after groups as a collective comprising both nurses and surgeons. Consequently, the impact of the recommendations was not individually evaluated for nurses or surgeons. The rationale behind this approach was rooted in recognizing the interdisciplinary nature of the problem. The intention was to gauge the overall teamwork performance rather than isolating the effects on nurses or surgeons separately.

A team comprising four surgeons and 23 nurses was established, and their performances were documented over a one-month period (28 days) before any recommendations were communicated to either group. Subsequently, the recommendations were implemented for a two-month period (56 days) involving the same team of surgeons and nurses. Notable improvements in surgical performance and success rates are evident in the table below. Due to unequal variance and a difference in the number of samples before and after implementation, Welch's test was conducted. The test details are provided in the table B-3, and it is worth noting that this test, different from the one performed in the published article, is more suitable for addressing this specific type of problem.

Metric	APT	APT_IN_ROOM	SPT	CASE	SFT	AFT	Turnover	**Success**
Mean-Before	40:54	16:28	12:07	1:03:36	04:50	05:44	15:00	**14%**
Mean-After	36:48	12:28	12:40	59:51	05:12	07:21	18:51	**64%**
STD-Before	13:20	09:11	06:04	08:22	01:59	03:05	33:13	0.3499
STD-After	11:08	10:01	07:21	12:18	02:20	03:52	49:06	0.229592
t-stat	1.3858	1.8347	-0.2209	1.6845	-1.4749	-2.3791	-0.3872	-690.9356
p-value	0.0862	**0.0358**	0.4129	**0.0482**	0.0723	**0.0102**	0.3499	**0.0000**
DF	46	58	63	72	71	61	74	37

Table B-3: Walsh test for the PD seminar recommendation implementation

4.3. Fundamental research contributions

First experiences of shaping organizational behavior via AI recommendations by positive deviance seminars.

4.4. Author Contributions

Contribution	Authors
Designed the study	R. Gold, F. Al Zoubi, D. Garvin, D. Schramm and A. Seely
Acquired the data	J. Brillinger and C. Kreviazuk
Wrote the manuscript	P. Fallavollita and P. Beaulé analyzed. R. Gold, F. Al Zoubi and D. Garvin
Critically revised	J. Brillinger, C. Kreviazuk, D. Schramm, P. Fallavollita, A. Seely and P. Beaulé
Final approval of the article to be published.	All authors

4.5. Article

The following article was accepted and published by the Canadian Journal of Surgery.

RESEARCH • RECHERCHE

Use of multidisciplinary positive deviance seminars to improve efficiency in a high-volume arthroplasty practice: a pilot study

Richard Gold, MD, MSc
Farid Al Zoubi, MSc
Julia Brillinger, MPH
Cheryl Kreviazuk, BA
Dennis Garvin, BSc, MBA
David Schramm, SM
Pascal Fallavollita, BEng, PhD
Andrew J.E. Seely, MD, PhD
Paul E. Beaulé, MD

Accepted Mar. 14, 2022

Correspondence to:
P. Beaulé
Division of Orthopaedic Surgery
Room W1640, Ottawa Hospital Research Institute
Box 502
501 Smyth Rd
Ottawa ON K1H 8L6
pbeaule@toh.ca

Cite as: *Can J Surg* 2023 January 3; 66(1). doi: 10.1503/cjs.018121

Background: Positive deviance (PD) seminars, which have shown excellent results in improving the quality of surgical practices, use individual performance feedback to identify team members who outperform their peers; the strategies from those with exemplary performance are used to improve team members' practices. Our study aimed to use the PD approach with arthroplasty surgeons and nurses to identify multidisciplinary strategies and recommendations to improve operating room (OR) efficiency.

Methods: We recruited 5 surgeons who performed high-volume primary arthroplasty and had participated in 4-joint rooms since 2012, and 29 nurses who had participated in 4-joint rooms and in at least 16 cases in our data set. Three 1-hour PD sessions were held in February and March 2021: 1 with surgeons, 1 with nurses, and 1 with both surgeons and nurses to select recommendations for implementation. The sessions were led by a member of the nonorthopedic surgical faculty who was familiar with the subjects discussed and with PD seminars. To determine the success of the recommendations, we compared OR efficiency before and after implementation. We defined success as performance of 4 joint procedures within 8 hours.

Results: Eleven recommendations were recorded from the session with nurses and 7 from the session with surgeons, of which 11 were selected for implementation. During the month after implementation, there were great improvements across all time intervals of surgical procedures, with the greatest improvements seen in mean anesthesia preparation time in the room (4.51 min [26.3%]), mean procedure duration (9.75 min [14.0%]) and mean anesthesia finish time (5.78 min [44.0%]) (all $p < 0.001$). The total time saved per day was 49.84 minutes; this led to a success rate of 69.0%, a relative increase of 73.8% from our 2012–2020 success rate of 39.7% ($p < 0.001$).

Conclusion: The recommendations and increased motivation owing to the individualized feedback reduced time spent per case, allowing more days to finish on time. Positive deviance seminars offer an inexpensive, efficient and collegial means for process improvement in the OR.

Contexte : Les séminaires de déviance positive (DP), une approche qui a déjà donné d'excellents résultats en termes d'amélioration de la qualité des pratiques chirurgicales, recourent à la rétroaction sur le rendement individuel pour identifier les membres des équipes dont le rendement excède celui de leurs pairs; les stratégies associées à tout rendement exemplaire servent à améliorer les pratiques des membres des équipes. Notre étude visait à utiliser la DP comme approche pour les chirurgiennes et chirurgiens et le personnel infirmier spécialisés en arthroplastie afin d'identifier des stratégies et des recommandations multidisciplinaires pour améliorer l'efficience des blocs opératoires (BO).

Méthodes : Nous avons recruté 5 spécialistes dont le volume d'interventions pour arthroplastie primaire était élevé et qui œuvraient dans des blocs à 4 interventions chirurgicales depuis 2012, et 29 membres du personnel infirmier ayant participé à la même cadence d'interventions et à au moins 16 cas de notre ensemble de données. Trois séances d'une heure ont eu lieu en février et mars 2021 : 1 avec les chirurgiens, 1 avec le personnel infirmier et 1 avec les 2 équipes pour choisir les recommandations à mettre en œuvre. Les séances étaient animées par un membre de la Faculté de chirurgie (non orthopédique) qui connaissait les sujets abordés et les séminaires de DP. Pour déterminer la réussite des recommandations, nous avons comparé l'efficience des BO avant et après leur mise en œuvre. La réussite se définissait par la réalisation de 4 arthroplasties en 8 heures.

Résultats : Onze recommandations ont été dégagées de la séance avec le personnel infirmier et 7 de la séance avec les chirurgiennes et chirurgiens; 11 ont été retenues en vue de leur application. Durant le mois suivant leur mise en œuvre, la durée des interventions s'est grandement améliorée; les améliorations les plus marquantes concernaient la durée moyenne de la préparation de l'anesthésie au bloc opératoire (4,51 min [26,3 %]), la durée moyenne des interventions (9,75 min [14,0 %]) et le temps de réveil moyen (5,78 min [44,0 %]) (tous $p < 0,001$). Le temps total gagné quotidiennement a été de 49,84 min; le taux de réussite a donc été évalué à 69,0 %, correspondant à une augmentation relative de 73,8 % par rapport à notre taux de réussite en 2012–2020 de 39,7 % ($p < 0,001$).

Conclusion : Les recommandations et la motivation accrue découlant de l'exercice de rétroaction individualisée a réduit le temps requis pour chaque cas et a permis de terminer plus de journées à temps. Les séminaires de DP sont une façon peu coûteuse, efficiente et collégiale d'améliorer les procédés au bloc opératoire.

Delivery of health care is coming to an inflection point in regard to supply and demand, with joint replacement demand in Canada increasing from 2014 to 2019 by 20.1% for hip replacement and 22.5% for knee replacement.[1] Increasing demand, combined with delays related to the COVID-19 pandemic, has created a large backlog of surgical procedures, especially in countries with universal health care like Canada and the United Kingdom, where efficient delivery of health care is even more critical.[2–4] Prolonged surgical wait-lists are further compounded by systemic inefficiencies: in North America, health care functions at a productivity level of about 43%; in the surgical care setting, inefficient use of time and space accounts for 30% of costs.[5]

To address this, various initiatives to increase throughput, such as high-efficiency operating rooms (ORs) and parallel processing with anesthesia block rooms, have been suggested.[6] At our institution, to address wait times and increasing demands, 4-joint rooms were instituted in 2004, but successful completion of 4 joint replacement procedures within the assigned OR time (i.e., 4 joints between 0730 and 1530) has been inconsistent.[7] This lack of efficiency, with overtime and lack of improvement, can lead to staff disengagement, fatigue and a sense of impossibility of the task at hand.[8]

To foster self-improvement and staff engagement to work as a team, various models of team efficiency have been developed using the Lean method, Six Sigma and process mapping, which can be quite effective but very resource intensive.[9,10] An alternative approach that has shown excellent results in improving the quality of individual surgeon practices is positive deviance (PD) seminars,[11,12] which use individual performance feedback to identify team members who outperform their peers. The strategies from those with exemplary performance are used to both motivate peers and improve the practices for all. Positive deviance has been effectively used in health care, public health, education and the private sector.[13] Positive deviance seminars focus on individual strengths and resources already present, instead of negatives that require improvement. Implementing the strategies is feasible and sustainable, as they are already in place and successful.

To our knowledge, PD seminars have not been studied in a multidisciplinary setting to improve OR performance and efficiency. Our study aimed to use the PD approach with arthroplasty surgeons and OR nurses to identify multidisciplinary strategies and recommendations to improve OR efficiency in running a single room to perform 4 primary joint replacement procedures within an 8-hour window.

METHODS

Study design and setting

Three 1-hour PD sessions were held in February and March 2021: 1 with surgeons, 1 with nurses, and 1 with both surgeons and nurses to select the recommendations for implementation. We recruited 5 surgeons, including P.E.B., who performed high-volume primary arthroplasty and had participated in 4-joint rooms since 2012. The 29 nurses selected for the study were those who had participated in 4-joint rooms and in at least 16 cases in our data set (prospectively recorded surgical records). The sessions were led by a member of the nonorthopedic surgical faculty (A.J.E.S.) who was familiar with the subjects discussed and with PD seminars to allow for a moderated discussion and avoid conflict.

Performance based on historical data for finishing on time (i.e., by 1530) as well as specific time intervals were shared with the respective individuals. The sessions first identified the PD strategies using provider-specific reports. Afterward, a confidential group interview gave all participants the opportunity to discuss the strategies they used that led to their success. At the end of the sessions, all the recommendations were voted on, and those with a unanimous vote were selected for implementation. The recommendations were implemented in the week after the last PD session.

Measures

The time interval data were recorded by the circulating nurse using the Surgical Information Systems. In addition, patient demographic characteristics (age, gender, body mass index, American Society of Anesthesiologists class), adverse events, nurse, anesthesiologist, 90-day readmissions and type of anesthetic used were collected. The time intervals used to determine OR efficiency were a modified version of those defined by the Association of Anesthesia Clinical Directors:[14] anesthesia preparation time; patient in room to anesthesia ready, surgical preparation time; anesthesia ready to procedure start, procedure duration; (procedure start time to procedure finish), anesthesia finish time; procedure finish to patient out of room, and turnover time; start of room cleanup to patient in room. The anesthesia preparation time immediately follows turnover time, as no delays are expected once the room is ready for the next case.

After the recommendations were implemented, the surgeons and nurses completed a survey on each 4-joint day to determine compliance with the recommendations and give feedback on their perception of the usefulness of the recommendations.

Statistical analysis

To determine the success of the PD sessions, we tested the strategies as a hypothesis in the prospective cohort. We determined whether the strategies improved time efficiency by comparing the time intervals observed in the prospective cohort to those previously observed. We used a paired *t* test to compare differences in mean interval times and rate of success in finishing on time between the pre- and postintervention groups.

Ethics approval

The Ottawa Health Science Network Research Ethics Board granted an ethics review exemption given that the project was considered to fall under continuous quality improvement rather than human subject research, as per Canada's Tri-Council Policy Statement on Ethical Conduct for Research Involving Humans.[15]

RESULTS

A total of 11 recommendations were recorded from the nurses' session and 7 from the surgeons' session (Box 1). Of the 18 recommendations, 11 were selected for implementation based on the combined session with nurses and surgeons:

- Recommendation 1: As a team, take time at the beginning of the day (at 0745, led by orthopedic staff or fellow) to go over the basics of the day's 4 cases.
- Recommendation 2: For 4-joint days, have a nursing team familiar with 4-joint rooms who have received total joint training. The team lead (care facilitator) should have adequate time for registered nurse training and administration (mechanism to be determined).
- Recommendation 3: Surgeon should be there for each case, from positioning to patient transfer from the table.
- Recommendation 4: Adopt a standardized approach for each type of procedure.
- Recommendation 5: Surgeon to review all 4 cases in advance to avoid last-minute changes on the day of surgery that require additional instruments that were not ordered.
- Recommendation 6: Bring the patient into the room for spinal preparation such that instruments may be opened simultaneously in parallel rather than in series with anesthesia preparation. However, ensure surgery and anesthesia have reasonable expectations of nurses during this period and have communicated with nurses.
- Recommendation 7: Anticipate next surgical steps. Call for instruments/implants in advance (and ensure fellows are doing this as well).
- Recommendation 8: Augment and facilitate putting away instruments during closure in the most efficient manner possible. Surgeons are encouraged to offer assistance with putting away instruments in coordination with the nursing staff.
- Recommendation 9: Change culture of multiple telephone calls to surgical team in the OR with questions. Have fellows contact the pre- and postanesthesia care units between cases to minimize need for telephone calls during cases.

Box 1. Suggested recommendations from sessions with surgeons and nurses

Surgeons

- Be there from positioning to patient transfer from the table.
- Have a standardized/protocolized approach for each type of procedure.
- Anticipate next steps, calling for instruments/implants.
- Assist with turnover and putting away instruments, but in a way that is supported by nurses.
- Institute an incentivization for the entire team to be done by 1530, and that would drive efficiency.
- Bring the patient into the room for spinal preparation such that instruments may be opened simultaneously (in parallel rather than in series).
- Anesthesia does the blocks and spinals in the procedure room.

Nurses

- Have an engaged, familiar team working together.
- Have equipment ready to go before the patient enters the room.
- Whole team (nursing, surgery, anesthesia) is present during turnover.
- Begin putting away instrumentation during closing.
- Have experienced, knowledgeable scrub nurses who know the steps to the procedure and will know when certain instruments (implants) are needed.
- Have attendants available to help with turnover.
- Ensure nurses in the room have received total-joint training.
- Minimize telephone call interruptions from the pre- and postanesthesia care units during the case.
- Ensure attending available for prepping. Make use of free staff in room when prepping/positioning. Ensure no revision of surgical positioning.
- Team lead must have adequate time for training and administration.
- Ensure improvements in efficiency do not come at the cost of patient outcomes.

- Recommendation 10: Have the whole team, including OR attendants, present in OR during turnover.
- Recommendation 11: Attendants should close communication loop with nursing team to ensure equipment (e.g., general equipment in the room, suction available, correct bed, extension cord for the bed) is ready to go at the start of each case.

During the month after the recommendations were implemented, there were great improvements across all time intervals of surgical procedures (Table 1). The greatest improvements were seen in mean anesthesia preparation time in the room (4.51 min [26.3%]), mean procedure duration (9.75 min [14.0%]) and mean anesthesia finish time (5.78 min [44.0%]) (all $p < 0.001$). Although each improvement seems small, the improvements totalled an average of 12.48 minutes per case, or 49.84 minutes per day. The total time saved was substantial, as it led to a success rate of 69.0%, a relative increase of 73.8% from our 2012–2020 success rate of 39.7% ($p < 0.001$).

Compliance and feedback

All 5 surgeons and 18 nurses completed the survey. The recommendations were all well received by the participants.

Recommendation 1

Team members did not follow this recommendation consistently. The only feedback was that this recommendation is redundant when the same operations are being performed consistently.

> This really doesn't impact on a 4–primary-joint day, as all team members know what is being done as we do over a thousand of these each year. So, meeting to say we are doing 4 primary joints is a bit redundant in my opinion and won't make the day run any faster. (Surgeon [success])

> Not for the 4 cases but the nurses know what to do when we have to do 4 joints with [the surgeon for these cases]. (Nurse [success])

Recommendation 2

All participants agreed that having nursing staff members who are familiar with the intricacies of total joint replacement was highly valuable. Participants indicated that it facilitated better teamwork in the OR, which led to everyone feeling less rushed.

Recommendation 3

Having the surgeon present in the room while the patient was being positioned was appreciated by nursing. On a day when the nurse had to call the surgeon for every patient to be positioned, the day did not finish on time.

Recommendation 4

Recently at our centre, as part of another initiative, we reduced the number of instrument trays to speed up the process in the OR and reduce waste (SLIM study[16]). Nursing felt this initiative improved efficiency in the OR substantially. In addition, on days on which all 4 patients were positioned the same (i.e., lateral decubitus), it simplified preparing for the case and positioning the patient.

Recommendation 5

Participants indicated that, on days on which cases were not reviewed in advance or were not relayed to the whole team, the day did not finish on time. Last-minute changes led to having the wrong equipment available, which led to delay in retrieving and opening a new equipment tray.

Recommendation 6

The most common feedback was regarding parallel processing; this was mentioned 11 times in the comments. On days on which parallel processing occurred, participants felt less rushed, and the days consistently finished early. Even on days with unexpected delays, parallel processing allowed for the day to finish on time. An issue with parallel processing is that anesthesia and nursing need to work as a team to allow this process to happen. On 1 instance, a lack of communication with anesthesia and nursing led to delays in patients' entering the room. There was no

Table 1. Observed changes in operating room performance after positive deviance sessions*

	Mean		Variance							
Facet of performance	After	Before	After	Before	Pooled variance	Hypothesized mean difference	df	*t* stat	*p* (*T* ≤ *t*) 1-tail†	*t* critical 1-tail
Success rate, %	0.65	0.40	0.23	0.24	0.24	0	792	3.74	**9.83**	1.65
APT, min	36.51	39.33	129.08	414.74	394.54	0	792	−1.03	0.15	1.65
APT in room, min	12.42	16.28	100.29	56.92	59.99	0	792	−3.63	**0.00**	1.65
SPT, min	12.63	14.39	54.12	28.79	30.58	0	792	−2.32	**0.01**	1.65
Time per case, min	59.62	69.33	154.36	296.15	286.12	0	792	−4.17	**1.67**	1.65
SFT, min	5.15	5.48	5.58	16.30	15.54	0	792	−0.62	0.27	1.65
AFT, min	7.30	13.25	15.08	50.15	47.67	0	792	−6.27	**2.99**	1.65
Turnover, min	26.64	24.21	63.09	69.60	69.14	0	792	2.12	**0.02**	1.65

AFT = anesthesia finish time; APT = anesthesia preparation time; df = degrees of freedom; SFT = surgical finish time; SPT = surgical preparation time.
*Number of observations = 57 after, 737 before.
†Bolded numbers indicate significant values.

coordination between the anesthesia procedure room and the OR; thus, the patient would wait for some time outside the OR, which led to delays and potential issues with the spinal dosage. In addition, 1 nurse mentioned the difficulty in properly teaching orientees while assisting with anesthesia with induction.

On days on which parallel processing did not occur, the surgeons felt more rushed. There were more perceived delays and idle time between cases.

> I had the most efficient anesthetist at our campus, and my OR surgical times from start to finish were all under an hour, at 53, 59, 56 and 54 minutes for the 4 cases. With all of that, we ended the day at 3:25 pm, leaving a whole 5 minutes to spare!! (Surgeon [success])

Recommendation 7

Participants indicated that next surgical steps were anticipated on most occasions. There was no feedback specific to this recommendation.

Recommendation 8

Participants noted that surgeons always assisted with cleaning up after cases. Although the assistance of surgeons in cleanup was appreciated, the fewer and slimmed-down instrument trays reduced the requirement for additional assistance for cleanup.

Recommendation 9

Participants indicated that the need for telephone calls during cases was minimized on most occasions; on other occasions, they were unsure.

Recommendation 10

Participants noted that the whole team, including OR attendants, were present in the OR during turnover 3 out of 8 times (38%). On 2 of the days when the whole team were not present, the day finished late. Several participants mentioned great appreciation for the presence of more nurses.

> Having a minimum of 3 nurses (2 circulating, 1 scrub) is essential. (Nurse [success])

> It was nice to have 2 RNs and 1 RPN scheduled in the 4-joint day. (Nurse [success])

Recommendation 11

Participants indicated that attendants always closed the communication loop with the nursing team to ensure that equipment was ready to go at the start of each case.

DISCUSSION

Positive deviance seminars are a simple yet effective and inexpensive means to facilitate collaboration across multiple disciplines in order to improve a process by learning by top performers, in our case, a high-volume arthroplasty practice. Our pilot study showed promising improvements in efficiency in the OR with the use of recommendations derived from PD seminars. Every stage of the surgical procedure except turnover showed substantial improvements compared to our mean performance in the previous 8 years. In addition to generating recommendations, the seminars gave participants the opportunity to review and reflect on their performance. Improved efficiency is not only relevant financially but also shortens wait times, since longer wait times are associated with prolonged recovery and decreased patient satisfaction,[17,18] as well as poor health care provider experience.[8,19]

Quality delivery of care involves many processes and can be assessed and improved through 3 sequential and interrelated dimensions: structures, processes and outcomes.[20] As health care is always evolving, with advances in technology, procedures and medical knowledge, the need for a “plan, do, study, act” cycle is necessary whereby data can be analyzed and feedback provided constantly in order to continuously revise and improve.[21] The plan, do, study, act cycle forms the basis for continuous quality improvement, which encompasses processes associated with providing a health care outcome.[22] In this study, the PD approach fit well as a continuous quality-improvement initiative in which we structured a planning approach to evaluate current structures and processes of joint arthroplasty care in the OR to improve them and thus achieve the desired outcome and vision, i.e., finishing within an 8-hour OR shift.

Many of the current best-practice articles in the OR efficiency literature involve a thorough process analysis, which entails mapping the entire process of the OR, from the patient entering the hospital to exiting the hospital.[9,10,23] Unlike those initiatives, PD focuses on the positive outliers: what is being done optimally, not what could be improved. In merely 3 hour-long sessions, participants were able to derive effective strategies, which would have otherwise taken many months if being derived by analyzing the complete process. All the new measures were taken from practices and resources already in place, which made for simple implementation: the measures were implemented within a week after the seminars were completed. In addition, implementing solutions that are not already in place can create unintended consequences, and the solutions are often unsustainable in the long term.[13]

When surgeons are not involved with various stages of a study, they often choose not to participate.[24] Positive deviance seminars gave team members the opportunity to collaborate in many stages of this study, including data analysis, by determining who the best performers were, and discussing and creating new recommendations. With team members engaged in multiple stages of the study, almost all were enthusiastic to participate and improve their practices. Positive deviance works particularly well with surgeons, as they are independent in practice and decision-making. Kim and Choi[25] identified that the most important factor in discovering innovative behaviours is the degree of independence that individuals have. Having independence allows employees to accept social risk by standing for their own ideas and disagreeing with coworkers in a positive manner.

We have previously reported that success rates of 4-joint rooms vary among surgeons, anesthesiologists and nurses.[7] By reviewing the practices of their peers, the participants in that study were able to identify gaps in their own practices and improve in the areas requiring focus. In addition, the present study shows that quantifying performance leads to improved motivation, as participants can understand how the time intervals of their practices compare to those of their peers and which areas of their practices require improvement. For example, a surgeon may perform a procedure faster than their peers but have the slowest positioning time. Identifying this issue might lead to the surgeon's placing more focus on ensuring that they are present in the OR during that time.

Although the morning huddle may feel redundant, its effectiveness is evident. Wright and colleagues[23] found that a common reason for delay of the first case was surgeon and anesthesiologist unavailability. By implementing a morning huddle, they were able increase the rate of on-time starts from 53% to 69%.

Limitations

This study has several limitations. It was carried out during the third wave of the COVID-19 pandemic, which led to cancellation of elective surgical procedures 1 month after the recommendations were implemented. In addition, during the monitoring period, the unique constraints of the pandemic may have led to a surgical environment that was not necessarily comparable to that before the pandemic. Some of the perioperative disciplines declined to participate in the seminars; full involvement of all disciplines would have improved the impact of the PD initiatives.

As in any study, the short time frame of monitoring is subject to confounding by the Hawthorne effect (a type of reactivity in which individuals modify an aspect of their behaviour in response to their awareness of being observed). As the nature of PD seminars does not allow for blinding, our studies could have not been designed in a way in which the participants were unaware of monitoring. However, this constant awareness likely contributed to increased motivation and effort. When we examined the results by individual surgeon, only 2 of the 5 surgeons showed an improvement in their success rates (64% to 81%, and 8% to 71%, respectively). Two surgeons did not perform any 4-joint rooms during the monitoring period, and 1 surgeon did not have any successful days.

One nurse shared that they felt rushed and that implementing all recommendations was unsustainable and too physically demanding. However, the overall consensus was that the recommendations made for a better team environment. Given the short duration of the monitoring period, a follow-up audit of OR efficiency would determine the sustainability and effectiveness of the recommendations.

CONCLUSION

The recommendations generated by the participating nurses and surgeons in the PD seminars, together with increased motivation owing to the self-performance feedback, reduced time per operation significantly, which increased the daily success rate. Positive deviance seminars offer an inexpensive, efficient, collegial and positive means for process improvement in the OR setting.

Affiliations: From the Division of Orthopaedic Surgery, The Ottawa Hospital, Ottawa, Ont. (Gold, Brillinger, Kreviazuk, Beaulé); the Faculty of Medicine, McGill University, Montréal, Que. (Gold); the Interdisciplinary School of Health Sciences, University of Ottawa, Ottawa, Ont. (Al Zoubi, Fallavollita); the School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, Ont. (Al Zoubi, Fallavollita); Clinical Operations, The Ottawa Hospital, Ottawa, Ont. (Garvin); the Department of Otolaryngology — Head and Neck Surgery, The Ottawa Hospital, Ottawa, Ont. (Schramm); and the Department of Thoracic Surgery, The Ottawa Hospital, Ottawa, Ont. (Seely).

Competing interests: None declared.

Contributors: R. Gold, F. Al Zoubi, D. Garvin, D. Schramm and A. Seely designed the study. J. Brillinger and C. Kreviazuk acquired the data, which P. Fallavollita and P. Beaulé analyzed. R. Gold, F. Al Zoubi and D. Garvin wrote the manuscript, which J. Brillinger, C. Kreviazuk, D. Schramm, P. Fallavollita, A. Seely and P. Beaulé critically revised. All authors gave final approval of the article to be published.

References

1. *Hip and knee replacements in Canada: CJRR annual statistics summary, 2018–2019*. Ottawa: Canadian Institute for Health Information; 2020.
2. Cram P, Landon BE, Matelski J, et al. Hip and knee arthroplasty utilization and outcomes in the United States and Canada: an analysis of New York and Ontario administrative data. *Arthritis Rheumatol* 2018;70:547-54.
3. OMA estimates pandemic backlog of almost 16 million health-care services [news release]. Toronto: Ontario Medical Association; 2021 June 9. Available: https://www.oma.org/newsroom/news/2021/jun/oma-estimates-pandemic-backlog-of-almost-16-million-health-care-services/ (accessed 2021 July 17).
4. Surgeons call for a 'New Deal for Surgery' to reduce the 'colossal' elective backlog [news release]. London (UK): Royal College of Surgeons of England; 2021 May 28. Available: https://www.rcseng.ac.uk/news-and-events/media-centre/press-releases/new-deal-for-surgery-2021/(accessed 2021 July 17).

5. Fairley M, Scheinker D, Brandeau ML. Improving the efficiency of the operating room environment with an optimization and machine learning model. *Health Care Manag Sci* 2019;22:756-67.
6. Waly FJ, Garbuz DS, Greidanus NV, et al. Safety of a "swing room" surgery model at a high-volume hip and knee arthroplasty centre. *Bone Joint J* 2020;102-B:112-5.
7. Beaulé PE, Frombach AA, Ryu JJ. Working toward benchmarks in orthopedic OR efficiency for joint replacement surgery in an academic centre. *Can J Surg* 2015;58:408-13.
8. Greenglass ER, Burke RJ, Fiksenbaum L. Workload and burnout in nurses. *J Community Appl Soc Psychol* 2001;11:211-5.
9. Cima RR, Brown MJ, Hebl JR, et al.; Surgical Process Improvement Team, Mayo Clinic, Rochester. Use of Lean and Six Sigma methodology to improve operating room efficiency in a high-volume tertiary-care academic medical center. *J Am Coll Surg* 2011;213:83-92, discussion 93-4.
10. Attarian DE, Wahl JE, Wellman SS, et al. Developing a high-efficiency operating room for total joint arthroplasty in an academic setting. *Clin Orthop Relat Res* 2013;471:1832-6.
11. Ivanovic J, Anstee C, Ramsay T, et al. Using surgeon-specific outcome reports and positive deviance for continuous quality improvement. *Ann Thorac Surg* 2015;100:1188-94, discussion 1194-5.
12. Ivanovic J, Mostofian F, Anstee C, et al. Impact of surgeon self-evaluation and positive deviance on postoperative adverse events after non-cardiac thoracic surgery. *J Healthc Qual* 2018;40:e62-70.
13. Baxter R, Taylor N, Kellar I, et al. What methods are used to apply positive deviance within healthcare organisations? A systematic review. *BMJ Qual Saf* 2016;25:190-201.
14. Boggs SD, Tsai MH, Urman RD; Association of Anesthesia Clinical Directors. The Association of Anesthesia Clinical Directors (AACD) Glossary of Times Used for Scheduling and Monitoring of Diagnostic and Therapeutic Procedures. *J Med Syst* 2018;42:171.
15. Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, and Social Sciences and Humanities Research Council. Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans. 2018. Available: https://ethics.gc.ca/eng/documents/tcps2-2018-en-interactive-final.pdf (accessed 2021 Apr. 15).
16. Adamczyk AP, Kim PR, Horton I, et al. The SLIM study: economic, energy, and waste savings through lowering of instrumentation mass in total hip arthroplasty. *J Arthroplasty* 2022;37(8S):S796-802.e2.
17. Hawker GA, Croxford R, Bierman AS, et al. All-cause mortality and serious cardiovascular events in people with hip and knee osteoarthritis: a population based cohort study. *PLoS One* 2014;9:e91286.
18. Tarride JE, Haq M, O'Reilly DJ, et al. The excess burden of osteoarthritis in the province of Ontario, Canada. *Arthritis Rheum* 2012;64:1153-61.
19. Bodenheimer T, Sinsky C. From triple to quadruple aim: care of the patient requires care of the provider. *Ann Fam Med* 2014;12:573-6.
20. Donabedian A. Evaluating the quality of medical care. 1966. *Milbank Q* 2005;83:691-729.
21. Su AW, Habermann EB, Thomsen KM, et al. Risk factors for 30-day unplanned readmission and major perioperative complications after spine fusion surgery in adults: a review of the National Surgical Quality Improvement Program database. *Spine* 2016;41:1523-34.
22. Beaulé PE, Roffey DM, Poitras S. Continuous quality improvement in orthopedic surgery: changes and implications with health system funding reform [editorial]. *Can J Surg* 2016;59:149-50.
23. Wright JG, Roche A, Khoury AE. Improving on-time surgical starts in an operating room. *Can J Surg* 2010;53:167-70.
24. Harvey EJ. Why don't my surgeries start on time? [editorial]. *Can J Surg* 2010;53:148-9.
25. Kim MJ, Choi JN. Group identity and positive deviance in work groups. *J Soc Psychol* 2018;158:730-43.

Chapter 5: Artificial-Intelligence driven prescriptive model to optimize team efficiency in a high-volume primary arthroplasty practice

5.1. Summary

In this chapter, we focused on a specific framework from the previous chapter, namely the Workflow Time Monitoring Framework (WTMF), and advanced it to a prescriptive model. The selection of this framework was based on various factors, including its flexibility and ease of implementation, as discussed in the previous chapter. In this journal, we established the benchmarks necessary to run the prescriptive model and identified key metrics for evaluation. Furthermore, we explained how a set of actions could be implemented to convert the predictive outputs into a prescriptive model, working in synergy to enhance the overall Surgical Success Rate (SSR).

5.2. Methodology

In the preceding chapters, our approach unfolded with the execution of descriptive analytics in Chapter two. Our primary objective was to gain an understanding of our data and evaluate the fundamental factors contributing to delays in arthroplasty. Following this analysis, we systematically categorized the data into three key segments: patient demographics metrics, time-related metrics, and surgical team metrics. This segmentation served two crucial purposes. Firstly, as elucidated in Chapter 3, it ensured the versatility of our solution, accommodating diverse data access scenarios within different institutions. For instance, not all institutions possess access to time metrics that meticulously record the various stages of surgery. This consideration becomes particularly relevant as we strive to provide a universal solution applicable across varied institutional settings. Secondly, as detailed in the same chapter, our categorization aligned with the findings from the feature importance analysis conducted using the Boruta method, i.e., Surgical team metrics are the most important while patients' demographics are the least. This alignment is extensively explained in the published article of this chapter, offering a comprehensive understanding of our data categorization strategy.

Building upon the categorization elucidated above, Chapter three introduced three distinct frameworks for a decision support systems infrastructure. Within each framework, we meticulously explained the components of each module, delineating their functionalities and illustrating the inputs and outputs they entail. Our focus, however, centered predominantly on the machine learning engine module of each framework. This emphasis was deliberate, aimed at providing concrete examples of ML models input and output for clarity. The intention was to facilitate a thorough understanding of the core predictive modeling aspect within each framework. However, we recognize the multifaceted nature of decision support systems, encompassing components such as the recommendation system, dashboard, databases, and other critical elements. While we did not extensively elaborate on these components for the majority of frameworks, we made a deliberate exception for the WTMF in the subsequent chapter.

Chapter four delved into the design and testing of the recommendation system for the WTMF, utilizing the positive deviance seminar. This approach not only introduced a unique and innovative aspect to our decision support system but also aimed at leveraging positive deviance within surgical teams for performance improvement. By incorporating recommendations derived from the positive deviants, we sought to enhance the efficiency and teamwork dynamics within the surgical setting. This approach was detailed with a focus on the WTMF, offering a practical example of how the recommendation system could be customized for a specific framework.

In this chapter, we undertook the task of contextualizing the information from the preceding chapters, amalgamating them into a theoretical and cohesive solution design for the complete PAS system centered around the WTMF. We elucidated how time-related data could be seamlessly integrated into the machine learning engine of the WTMF to generate baseline benchmarks. If these benchmarks are not met, the recommendation system is activated to facilitate the catching up with subsequent benchmarks and overcoming potential delays throughout the day. The benchmarks, coupled with the surgery timeline and the SSR, are envisioned to be visually presented in a dashboard. This dashboard serves as a comprehensive tool, allowing users to assess their speed and evaluate the feasibility of completing each surgery within the designated timeframe.

The realization of the design outlined in this chapter comes to fruition in Chapter 7, where the designed solution is implemented and put into practical action. The implementation phase involves a meticulous application of our proposed PAS system, incorporating the WTMF and its associated components. It is during this phase that we witness the translation of our theoretical models and frameworks into real-world scenarios. The challenges encountered, lessons learned, and the outcomes achieved during this implementation provide valuable insights into the viability and effectiveness of our proposed solution. This chapter serves as the bridge between theory and application, highlighting the tangible impact of our research on optimizing surgical workflows and improving overall efficiency in the healthcare setting.

5.3. Fundamental research contributions.

Development of the first AI-driven recommendations that contribute to overtime and delays at TOH.

Author	Contribution to the study
Farid Al Zoubi	Methodology, Formal analysis, Investigation, Software, Validation, Visualization, Writing original draft, Writing–review and editing
Richard Gold	Data curation, Investigation, Validation, Visualization, Writing–original draft, Writing–review and editing
Stéphane Poitras	Methodology, Resources, Writing–original draft, Writing–review and editing
Cheryl Kreviazuk	Data curation, Investigation, Writing–review and editing
Julia Brillinger	Data curation, Investigation, Writing–review and editing
Pascal Fallavollita	Conceptualization, Methodology, Project administration, Resources, Software, Writing–original draft, Writing–review and editing
Paul Beaulé	Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing–original draft, Writing– review and editing

5.4. Article

The following article was accepted and published by the International Orthopedics Journal.

International Orthopaedics (2023) 47:343–350
https://doi.org/10.1007/s00264-022-05475-1

ORIGINAL PAPER

Artificial intelligence-driven prescriptive model to optimize team efficiency in a high-volume primary arthroplasty practice

Farid Al Zoubi[1] · **Richard Gold**[2,3] · **Stéphane Poitras**[4] · **Cheryl Kreviazuk**[2,5] · **Julia Brillinger**[2,5] · **Pascal Fallavollita**[6] · **Paul E. Beaulé**[2,5]

Received: 29 April 2022 / Accepted: 8 June 2022 / Published online: 27 June 2022

Abstract

Purpose We aimed to improve OR efficiency using machine learning (ML) to find relevant metrics influencing surgery time success and team performance on efficiency to create a model which incorporated team, patient, and surgery-related factors.
Methods From 2012 to 2020, five surgeons, 44 nurses, and 152 anesthesiologists participated in 1199 four joint days (4796 cases): 1461 THA, 1496 TKA, 652 HR, 242 UKA, and 945 others. Patients were 2461f:2335 m; age, 64.1; BMI, 29.93; and ASA, 2.45. Surgical Success was defined as completing four joints within an eight hour shift using one OR. Time data was recorded prospectively using Surgical Information Management Systems. Hospital records provided team, patient demographics, adverse events, and anesthetic. Data mining identified patterns and relationships in higher dimensions. Predictive analytics used ML ranking algorithm to identify important metrics and created decision tree models for benchmarks and success probability.
Results Five variables predicted success: anaesthesia preparation time, surgical preparation time, time of procedure, anesthesia finish time, and type of joint replacement. The model determined success rate with accuracy of 72% and AUC = 0.72. Probability of success based on mean performance was 77–89% (mean-median) if APT 14–15 minutes, PT 68–70 minutes, AFT four to five minutes, and turnover 25–27 minutes. With the above benchmarks maintained, success rate was 59% if surgeon exceeded 71.5-minutes PT or 89% if 64-minutes procedure time or 66% when anesthesiologist spent 17–19.5 minutes on APT.
Conclusion AI-ML predicted OR success without increasing resources. Benchmarks track OR performance, demonstrate effects of strategic changes, guide decisions, and provide teamwork improvement opportunities.

Keywords Operating room efficiency · Artificial intelligence · Machine learning · Arthroplasty · Teamwork

---

✉ Paul E. Beaulé

Farid Al Zoubi
falzo100@uottawa.ca

Richard Gold
richard.gold3@mail.mcgill.ca

Stéphane Poitras
Stephane.Poitras@uottawa.ca

Cheryl Kreviazuk
ckreviazuk@ohri.ca

Julia Brillinger
jbrillinger@ohri.ca

Pascal Fallavollita
pfavallo@uottawa.ca

[1] Faculty of Electrical Engineering and Computer Science (EECS), University of Ottawa, Ottawa, Canada

[2] Division of Orthopaedic Surgery, The Ottawa Hospital, Room W1640, 501 Smyth Road, Box 502, Ottawa, ON K1H8L6, Canada

[3] Faculty of Medicine, McGill University, Montreal, Canada

[4] School of Rehabilitation Sciences, University of Ottawa, Ottawa, ON, Canada

[5] Clinical Epidemiology Program, The Ottawa Hospital Research Institute, Ottawa, Canada

[6] Interdisciplinary School of Health Sciences, University of Ottawa, Ottawa, Canada

Introduction

Currently, healthcare functions at a productivity level of 43% in North America [1]. The demand for care is rapidly increasing, with budgets increasingly difficult to control, and regulations becoming ever more uncertain and complex, this situation is not sustainable and becoming increasingly challenging in the advent of bundle care payments [1]. In Canadian orthopaedic surgery, from 2014 to 2019, arthroplasty output increased by 20.1% for hip replacement and 22.5% for knee replacement [2]. Meeting the growing demand of joint replacements and Canadian provincial wait time targets led to numerous initiatives such as government-funded quality-based procedure funding, high-efficiency four joint operating room (OR) days (i.e., 4 joints within 8 h 7:30am–3:30 pm), and swing rooms (overlapping) [3, 4]. When implementing quality improvement changes, it is critical that appropriate resources are allocated to ensure efficiency and sustainability. Our initial experience with four joint rooms at our institution was disappointing with only a 49% success rate [5]. Inefficient use of time and space accounts for 30% of the costs of surgical care [1]. Accurately selecting which factors improve OR efficiency requires understanding relationships between patient, team (surgeon, anesthesiologist, nursing), and environment.

The increased need for optimized healthcare delivery motivated us to address the optimization of hospital surgery productivity using artificial intelligence (AI) to assist in surgery scheduling and productivity using team-specific data on a targeted surgical intervention to maximize OR efficiency and enhance multidisciplinary care delivery. As traditional analytics typically describe relationships between two and four factors, machine learning (ML), a subset of AI, is needed. Not only can ML establish benchmarks using many variables, but it also continuously improves accuracy as additional data is collected.

In each surgery, there are many components influencing efficiency such as team, patient, type of surgery, type of anaesthesia, environment, and equipment. Traditional analytics and descriptive statistics have limitations in defining benchmarks using a multitude of variables. Through AI, we can study the interaction of factors, determine which factors affect success rate, and create a model which incorporates team, patient, and surgery-related factors to develop accurate benchmarks for success. Not only would such a model predict if the day is going to be successful, but it would also show which section(s) of surgical delivery of care are responsible for delays, while highlighting solutions to improve efficiency. This prescriptive dynamic modeling would permit analysis of past performances, predict the future, and suggest solutions. Suggested solutions to reduce time in targeted stages of surgery can be collected from surgery staff or could be from changes in procedures [6].

We aimed to perform advanced analytics (AA) and machine learning (ML) techniques to find the relevant metrics and benchmarks that influence surgery time success and examine influence of team member performance to improve OR efficiency to achieve a success rate of 80% for performing four joints within an eight hours shift.

Patients and methods

Data set

All joint replacements completed as part of a four-joint room at a tertiary care centre from 2012 to 2020 were reviewed. A four joint room was defined as a scheduled eight hour day (7:30am–3:30 pm) where four unilateral joint replacements were done by the same surgeon. The day is considered complete when the patient exits the operating room. We identified five surgeons, 44 circulating nurses, and 152 anesthesiologists participating in 1199 four joint days (4796 cases). The breakdown in procedures was 1461 total hip arthroplasties (THA), 1496 total knee arthroplasties (TKA), 652 hip resurfacing (HR), 242 unicompartmental knee arthroplasties (UKA), and 945 of other procedures such as a combination of the above (i.e., bilateral THA) or patellofemoral joint replacement (PFJ). The patient demographics were 2461 females, 2335 males with average age of 64.1 (range 17–99), mean BMI of 29.93 (range 17.1–51.4), and ASA of 2.45 (range 1–4). Surgical success was defined as completing four joints within an eight hour shift using one OR theatre (7:30am–3:30 pm).

Time data was recorded at time of surgery using Surgical Information Management Systems (SIMS). Hospital records provided surgical team composition, patient demographics (age, sex, BMI, ASA), adverse events, 90-day readmissions, and anaesthetic type (general, spinal, regional). Time intervals used are a modified version of those defined by the Association of Anesthesia Clinical Directors (AACD): anaesthesia preparation time (APT); patient in-room to anaesthesia ready, surgical preparation time (SPT); anaesthesia ready to procedure start, procedure; procedure start to procedure finish, anaesthesia finish time (AFT); procedure finish to patient out of room, and turnover time; and first patient exits to subsequent patient in room (Fig. 1). APT immediately follows turnover. [7]

Feature engineering

We started with 40 features collected from the SIMS (Fig. 2). We used time stamps to derive intervals. ASA and BMI had a covariance of 0.85; thus, only BMI was selected. For the model, we eliminated values with very few samples (less than 5%) such as Circulator 3, Circulator 4, 90-day readmission, and post-operative complications. Some metrics such as date were expanded; and day, month, and year were needed for analyses. The initial 40 features were refined

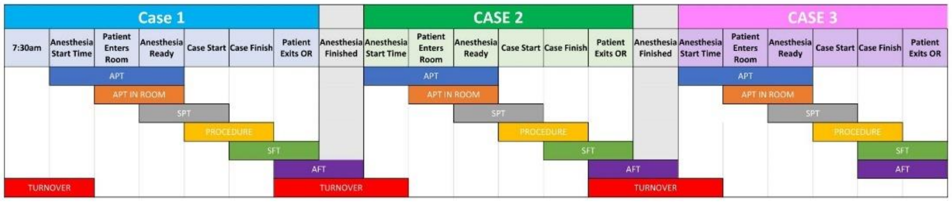


Fig. 1 Time intervals during a four joint day

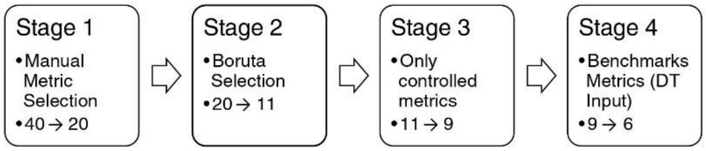


Fig. 2 Methodology of our feature selection criteria

to 20 for the Boruta analysis stage. Too many features may yield overfitting, where the model memorizes the data but cannot precisely predict outcomes [8] (Table 1).

Data for each metric was then prepared and cleaned according to the Cross-Industrial Standard Process for Data Mining (CRISP-DM) [9]. Categorical features with more than ten categories, such as anesthesiologist and nurse, had low-frequency values labeled as “other.” Low-frequency values were defined by normalizing the count of surgery and selecting individuals in the top 20% of cases completed. The different subtypes of anaesthesia were generalized into general, regional, or spinal.

Centricity metrics and other statistical metrics were plotted for time intervals used in benchmarks to determine the skewness of the data, kurtosis, standard deviation, and mean. Abnormal records were then eliminated (i.e., age – 1); skewness was handled by either taking the natural logarithmic of the feature like SFT and AFT or by binning to groups like age.

Descriptive analytics

The absolute relationship of each metric was analyzed independently along with the success rate of that metric. Success rate was calculated by dividing success number of surgeries by total number of surgical procedures, a success day would have four success records (rows), similarly for unsuccessful day, has four unsuccessful records. Other trends and relationships between multiple features, with up to four metrics at the same time, were also studied.

Machine learning

Four thousand seven hundred ninety-six surgical procedures were included in the data set. Seven hundred ninety-six records included all information about each case including patient (age, BMI, ASA, readmission), surgery (type, anesthesia, campus), team (surgeon, nurse, anesthesiologist), and

Table 1 List of initial features used for generating ML model

List of initial features used			
Surgeon	90-day readmissions	Case finish	Anesthesiologist 2 first name
Case no	Reason for readmission	Surgery finish time	Anesthesiologist 2 last name
Date	Length of stay	Turnover	Circulator RN 1 first name
Campus	Anaesthesia start time	Surgical time in minutes	Circulator RN 1 last name
Type	Time in room	Surgery	Circulator RN 2 first name
Type of anesthesia	Anaesthesia ready time	Total surgical time per day	Circulator RN 2 last name
Sex	Anaesthesia stop time	Time out of room	Circulator RN 3 first name
Age	Anaesthesia finish time	Turnover	Circulator RN 3 last name
BMI	Surgical preparation time	Anesthesiologist 1 first name	Circulator RN 4 first name
ASA	Case start	Anesthesiologist 1 last name	Circulator RN 4 last name

time (time stamps and date). The rest of the data set of 4000 surgical records did not include nurse or BMI.

The data set was used for the feature selection process using the Boruta algorithm, which reduced the input metrics to 13 [10]. Boruta categorizes metrics in order of importance. By removing peri-operative patient metrics like type of joint, type of anaesthesia, and age, the number of metrics were reduced to six: procedure, APT in room, SPT, surgery finish time, turnover, and AFT. These six metrics were then arranged from most important to least important using Boruta. All records were then used to delineate our decision tree model (Fig. 3).

Thereafter, a supervised machine learning classification algorithm was used which allowed prediction of the output (success/failure) and to receive the specific time intervals needed to achieve the said output. ML models that are interpretable are limited. Decision tree (DT) algorithm was selected as it explains the output in the desired format, benchmarks which use the decisions rules to describe the probability of each rule. DT models are best in our scenario as the relationship between dependent variables and outcomes is non-linear.

To verify our model and determine AUC-ROC, we used the sixfold cross-validation method where the data is divided into 6 equal subsets, which are all used for training and validation [11]. The output accuracy is then the average output of all six subsets. Cross validation handles the overfitting problem efficiently.

Results

Descriptive analytics

The mean success rate was 39% (annually ranging from 26 to58%). If a day finished late, the day would end on average 21 min late (range 0–60 min). The years 2013 and 2019 had the lowest success rates of 26% and 31%, respectively. Of note, in 2013, our institution changed implant vendors, and in 2019, we changed to a new electronic health record system. Success rate by surgeon ranged from 28 to 59%. Mean surgery time was 68.06 min (range between surgeons: 66.3–72 min). This trend was similarly observed for anesthesiologists and nurses.

Machine learning

Of the 40 metrics, 20 were initially selected to predict success: *team* (surgeon, nurse, anesthesiologist), *time intervals* (APT, SPT, procedure, AFT, surgery finish time, AFT), and *type of joint* were significant. Patient metrics such as gender, BMI, and ASA did not influence success. After removing uncontrollable metrics like patient age and type of surgery, we ended up having six main metrics as input to our DT model, for which we used for benchmarks (Fig. 4). Of the time intervals, procedure duration had the highest importance and the longest mean duration (69.6). However, turnover time had higher average duration 26.5 min (range 3–145 min) than SPT 8.8 min (range 0–75 min) and surgery finish time 5.2 min (range 0–49 min), yet it was ranked less important than both (Table 2).

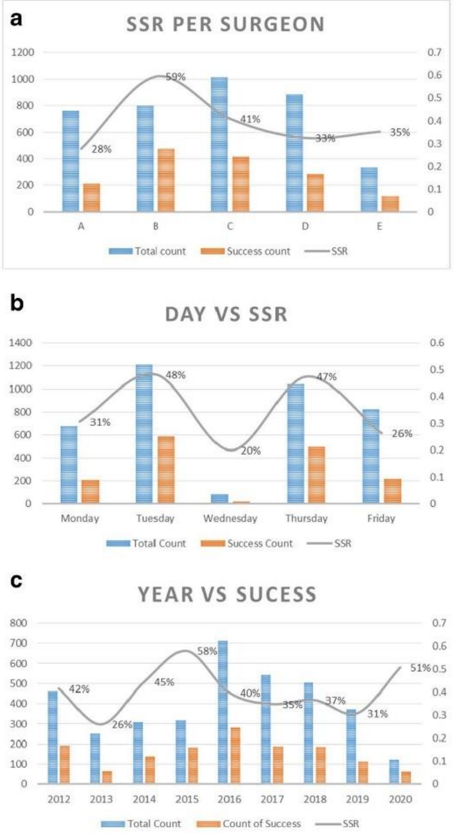


Fig. 3 (**a**) Success rate by surgeon from 2012 to 2020. (**b**) Success rate by day of the week from 2012 to 2020. (**c**) Success rate by year from 2012 to 2020

Model output

The predictive benchmark model determined success rate based on any time interval input of APT in room, procedure, surgery finish time, AFT, SPT, and turnover and calculated

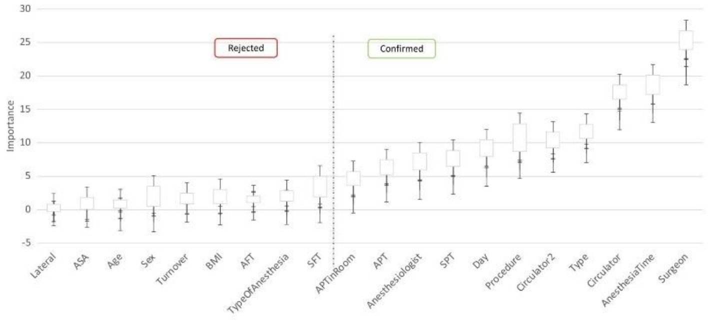


Fig. 4 Boruta output demonstrating the importance of the features selected for our AI model

Table 2 Distribution of the time intervals used for AI-Model

Metrics	SFT	APTInroom	SPT	Procedure	AFT	Turnover	Probability success
Min	0	0	0	19	0	3	100%
1st quartile	3	10	12	60	4	20	91%
Median	5	14	15	68	5	25	89%
Mean	5	15	15	70	7	27	77%
3rd quartile	6	18	17	77	8	31	0%
Max	49	68	75	210	74	145	0%

success with an accuracy of 72% (correctly predicted over total predicted) and AUC = 0.72 (Fig. 5). Our baseline probability, based on the mean performance of our team members from 2012 to 2020 (median to mean) was 77–89% if APT 14–15 min, procedure 68–70 min, AFT 4–5 min, and turnover 25–27 min. Team members have a large influence on success rate; with the above benchmarks maintained, success rate was 59% if surgeon exceeded 71.5-min procedure time or 89% if 64-min procedure time and 93% if 53-min procedure time or 66% when anesthesiologist spent 17–19.5 min on APT. However, if turnover reached 28 min, there was no change in probability of success.

Discussion

Artificial intelligence is widely used for clinical decision-making, diagnosis, and improving patient outcomes. IBM Watson focused on analyzing multidimensional patient data for decision making [12]. Google DeepMind Health used ML to analyze medical data for diagnosis and providing time-sensitive information for surgeons [13]. Both

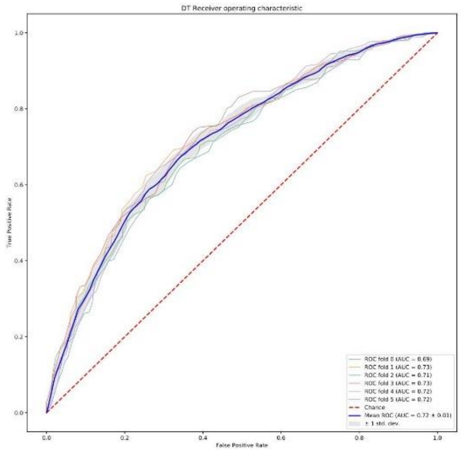


Fig. 5 Receiver operating characteristic of decision tree model

Intel Healthcare Analytics and Ascos CancerLinQ use Big Data analytics to improve patient therapy by analyzing patient electronic health records [14, 15]. Although there has been a large increase in the use of ML in healthcare, there has been a sparsity in the use of AI on throughput and organizational management [16]. ML has been used to predict OR durations and cases with a high risk of cancelation and to generate schedules to improve OR and PACU utilization rates [1, 17–19]. The above studies made use of random forest models to make accurate predictions. Unfortunately, random forest models do not have an interpretable output, making it difficult to implement meaningful changes [16].

Descriptive analytics shed light on our data and described the relationships of our features among each other and with success rate. However, our need was beyond merely studying the features one by one nor even few features together. We sought to demonstrate the impact of all metrics when assembled in a dynamic fashion that can describe the interactions of the important metrics and the impact of any change of any metrics independently or as whole. We aimed to determine which features are irrelevant and if a feature is important determining the ranking. We created an AI/ML model which can predict with an accuracy of 72% the success of a four joint room based on individual performance during each stage of surgery. The model creates a baseline to monitor individual performance and creates opportunities to make real-time strategic changes to improve the probability of success.

The Boruta model used for feature selection determined that composition of the team (surgeon, nurse, and anesthesiologist) was the most important for success. Our study showed that patient-related characteristics did not play a significantly influence success relative to the other factors which is similar to another study which denoted that patient factors explained only 25% of the variance in APT [20]. However, this differs from Strum et al. which concluded both surgeon and patient perioperative factors such as ASA, age, gender, and anesthesia type influenced surgical time [21]. However, unlike our study, they described which factors are or are not important, but did not quantify or rank relative importance [21]. Although the surgeon was the most important team member, all team members were significant drivers of success. In other words, if other team members do not perform, the likelihood of success falls significantly from 77 to 59–69% despite the surgeon performing well. Consequently, success cannot rest solely on the shoulders of the surgeon. A single team member underperforming can be the determining factor in a room finishing on time or not.

AI-generated benchmarks give team members multiple opportunities to achieve success. If a benchmark is not met, the model identifies which interval(s) of the surgical patient encounter can be used to catch up and how much time is required to be saved to finish on time. This is opposed to our earlier work on benchmarks which attempted to predict success of a four joint day but did not give team members opportunities to finish on time or give the probabilities of success of multiple benchmarks [5]. For example, if a case is suspected to take longer than usual by a team member (75 min), the AI-model has identified intervals (i.e., shorter APT) which when shortened can achieve almost the same success rate as our base case (74%). This provides an objective justification for reducing APT (8.5 min) by the anaesthesia team, which can be accomplished by performing induction outside the OR in a block room (Fig. 6).

Ultimately, the goal of this project was to improve OR throughput to reduce wait times and cost. A key finding of our model was one minute spent on the case is not equivalent to one minute spent on another interval; the relationships are not linear. For example, a 1.5-min increase in procedure time changed the success rate from 77 to 59%, whereas a two minute increase in turnover did not produce any. Thus, interventions used to improve success rates need to focus on improving the intervals which impact success rate. Some intervals have more room for improvement than others. Intervals such as surgery finish time with a mean of five minute (IQR 3–6 min) are short and therefore harder to compress or make improvements. However, APT can be an area for improvement because each minute had great impact on success rate (importance 9.8) and has a 25th to 75th quartile range of ten to 18 min, giving an eight minute window for improvement.

Referring to our model, maintaining the mean of all other intervals and achieving an APT in room of 10.5 min or less would have a success rate of 77%. This information can be quite helpful in determining impact of new processes such as an anesthesia procedure room. Achieving a success rate of 77% would be a 76% improvement compared to our base case. With our base case, 61% of four joint days end late, with an average of 36 min of overtime wages paid, which hypothetically, in a year with 250 four joint rooms that leads to 91.5 h of overtime and $312,051.60 in increased cost (overtime costs $56.84 per minute) [22]. If we can achieve a success rate of 77%, there will only be 34.5 hours of overtime and $117,658.80 in overtime wages, a saving of 57 h and $194,392.80 in costs. With both the time and money saved, the hospital will have increased time to possibly perform a fifth case, and the funding to pay for that case, reducing the burden of the joint replacement waitlist.

Benchmarks create the opportunity to monitor and quantify individual performance. Team members can take their past performances and use this information to understand which intervals of the surgery they can improve on. Individual performance feedback gives the opportunity to understand strengths and weaknesses in practices and create and

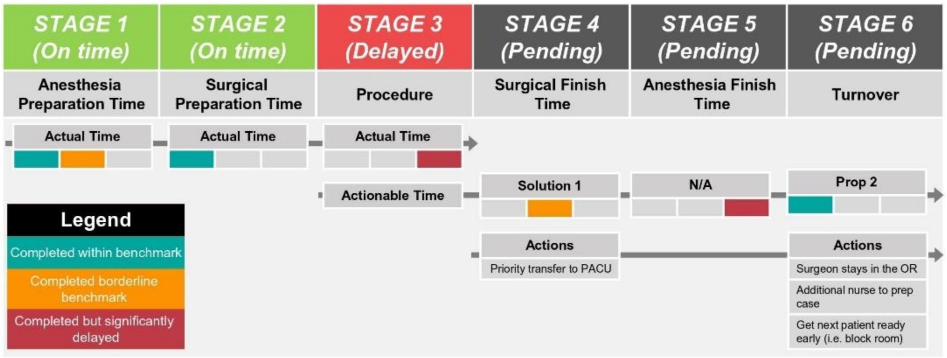


Fig. 6 Example of OR management dashboard implementing AI model

motivate improvement. Moreover, benchmarks create optimal frontiers for success in the OR setting. Thus, if there is a new nurse or anesthesiologist working that day, the surgeon will be aware of which interval of the surgery they can assist in or speed up to make sure the day finishes on time. In addition, this can be used as part of quality initiatives.

Wu et al. suggested using a cutoff benchmark, i.e., 75th or 90th percentile of worst performers and investigating the reasons why they failed, such as technical difficulties or inexperience which can be improved by further training [23]. Additionally, benchmarks can be used by management for hiring and evaluating performance. For example, during the probation period of a new staff member, the performance can be compared to the baseline to determine if they are a fit in the high-volume arthroplasty practice. By balancing out the patient perioperative factors when scheduling patients, we will be able to reduce variability in the case durations and prevent too many lengthy cases taking place in a single day.

Limitations

This study is not without its limitations. Initially, we felt a scheduling model with optimized teams would lead to the best results of time and cost savings. In one study, they used an AI model to schedule surgeries which improved robot-assisted surgery output by 34.9 to 51.7% [19]. With human resource constraints and workplace politics, a computer model to select teams would have benefited the high performers and negatively affected poor performers, which goes against the goal of our project to improve the performance of all. Using benchmarks instead of a scheduling platform allows the individual nurse, surgeon, and anesthesiologists to grasp what performance is expected of them to contribute the team effort. We did not want to only optimize human resource placement, but we also wanted to improve the team environment and OR performance of all. Other models which are more accurate but are not interpretable could have been selected. However, they do not detail the interaction and influence of individual intervals of the surgery. Although the model was only 72.8% accurate, this is acceptable. The overall accuracy of the model was less important than probability of success of each rule of the decision tree as this creates multiple opportunities to improve success. A machine learning model is limited by the dimensions of the data set matrix. In this study, we started with 40 features which were narrowed down to six. As time goes, inherent ambiguity of our machine learning model will decrease as the dimensions of our data set matrix will grow with both features and columns. In the future, behavioural data can be added to our data set matrix using blinded voice and video recordings to add more features to our model [24, 25].

Conclusion

AI successfully predicted OR success and generated benchmarks for arthroplasty. Benchmarks track OR performance, demonstrate effects of strategic changes, guide decisions, and provide opportunities to improve teamwork such as through positive deviance seminars. Predictive benchmarks can lead to more surgeries performed without increasing resources used, improving joint replacement wait times.

Author contribution Farid Al Zoubi: Methodology, Formal analysis, Investigation, Software, Validation, Visualization, Writing–original draft, Writing–review and editing

Richard Gold: Data curation, Investigation, Validation, Visualization, Writing–original draft, Writing–review and editing

Stéphane PoitrasMethodology, Resources, Writing–original draft, Writing–review and editing

Cheryl Kreviazuk: Data curation, Investigation, Writing–review and editing

Julia Brillinger: Data curation, Investigation, Writing–review and editing

Pascal FallavollitaConceptualization, Methodology, Project administration, Resources, Software, Writing–original draft, Writing–review and editing

Paul Beaulé: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing–original draft, Writing–review and editing

Funding This work was supported by the Ontario Ministry of Health.

Data availability The data that support the findings of this study are available from the corresponding author, Dr. Paul Beaulé, upon reasonable request.

Declarations

Ethics approval This project was reviewed and deemed as exempt from research by the Ottawa Health Science Research Ethics Board.

Consent to participate Not applicable

Consent to publish Not applicable

Conflict of interest The authors declare no competing interests.

References

1. Fairley M, Scheinker D, Brandeau ML (2019) Improving the efficiency of the operating room environment with an optimization and machine learning model. Health Care Manag Sci 22:756–767. https://doi.org/10.1007/s10729-018-9457-3
2. Canadian Institute for Health Information. (2020) Hip and knee replacements in Canada: CJRR Quick Stats, 2018–2019. Ottawa, ON
3. Cram P, Landon BE, Matelski J et al (2018) Hip and knee arthroplasty utilization and outcomes in the United States and Canada: an analysis of New York and Ontario administrative data. Arthritis Rheumatol (Hoboken, NJ) 70:547. https://doi.org/10.1002/ART.40407
4. FJ W, DS G, NV G, et al (2020) Safety of a "swing room" surgery model at a high-volume hip and knee arthroplasty centre. Bone Joint J 102-B:112–115. https://doi.org/10.1302/0301-620X.102B7.BJJ-2019-1536.R1
5. Beaulé PE, Frombach AA, Ryu J-J (2015) Working toward benchmarks in orthopedic OR efficiency for joint replacement surgery in an academic centre. Can J Surg 58:408. https://doi.org/10.1503/CJS.001215
6. Gold R, Al Zoubi F, Brillinger J, et al Use of multidisciplinary positive deviance seminars to improve efficiency in a high-volume arthroplasty practice: a pilot study. Can J Surg
7. SD B, MH T, RD U (2018) The Association of Anesthesia Clinical Directors (AACD) glossary of times used for scheduling and monitoring of diagnostic and therapeutic procedures. J Med Syst 42. https://doi.org/10.1007/S10916-018-1022-6
8. Ying X (2019) An overview of overfitting and its solutions. J Phys Conf Ser 1168:022022. https://doi.org/10.1088/1742-6596/1168/2/022022
9. Shearer C (2000) The CRISP-DM model: the new blueprint for data mining. J data Warehous 5:13–22
10. Kursa MB, Jankowski A, Rudnicki WR (2010) Boruta - a system for feature selection. Fundam Informaticae 101:271–285. https://doi.org/10.3233/FI-2010-288
11. Berrar D (2018) Cross-validation. Encycl Bioinforma Comput Biol ABC Bioinforma 1–3:542–545. https://doi.org/10.1016/B978-0-12-809633-8.20349-X
12. High R Front cover the era of cognitive systems: an inside look at IBM Watson and how it works
13. G A, W G, R M et al (2016) Hybrid computing using a neural network with dynamic external memory. Nature 538:471–476. https://doi.org/10.1038/NATURE20101
14. Raghupathi W, Raghupathi V (2014) Big data analytics in healthcare: promise and potential. Heal Inf Sci Syst 2https://doi.org/10.1186/2047-2501-2-3
15. GW S, RS M, R H (2013) CancerLinQ and the future of cancer care. Am Soc Clin Oncol Educ book Am Soc Clin Oncol Annu Meet 430–434. https://doi.org/10.14694/EDBOOK_AM.2013.33.430
16. Bellini V, Guzzon M, Bigliardi B, et al (2020) Artificial intelligence: a new tool in operating room management. Role of machine learning models in operating room optimization. J Med Syst 44 https://doi.org/10.1007/S10916-019-1512-1
17. Luo L, Zhang F, Yao Y, et al (2018) Machine learning for identification of surgeries with high risks of cancellation 26 141–155. https://doi.org/10.1177/1460458218813602
18. Shahabikargar Z, Khanna S, Sattar A, Lind J (2017) Improved prediction of procedure duration for elective surgery. Stud Health Technol Inform 239:133–138. https://doi.org/10.3233/978-1-61499-783-2-133
19. Zhao B, Waterman RS, Urman RD, Gabriel RA (2019) A machine learning approach to predicting case duration for robot-assisted surgery. J Med Syst 43:
20. Maheshwari K, You J, Cummings KC et al (2017) Attempted development of a tool to predict anesthesia preparation time from patient-related and procedure-related characteristics. Anesth Analg 125:580–592. https://doi.org/10.1213/ANE.0000000000002018
21. Dp S, Ar S, Jh M, LG V, (2000) Surgeon and type of anesthesia predict variability in surgical procedure times. Anesthesiology 92:1454–1466. https://doi.org/10.1097/00000542-200005000-00036
22. Petis S, Howard J, Lanting B, et al In-hospital cost analysis of total hip arthroplasty: does surgical approach matter? Elsevier
23. Wu HL, Chang WK, Hu KH, et al (2015) A quantile regression approach to estimating the distribution of anesthetic procedure time during induction. PLoS One 10. https://doi.org/10.1371/journal.pone.0134838
24. Jung JJ, Jüni P, Lebovic G, Grantcharov T (2020) First-year analysis of the operating room black box study. Ann Surg 271:122–127. https://doi.org/10.1097/SLA.0000000000002863
25. Mascagni P, Padoy N (2021) OR black box and surgical control tower: recording and streaming data and analytics to improve surgical care. J Visc Surg 158:S18–S25. https://doi.org/10.1016/J.JVISCSURG.2021.01.004

Chapter 6: Leveraging Machine Learning and Prescriptive Analytics to Improve Operating Room Throughput

6.1. Summary

In this chapter, the journal explores the potential utilization of overtime compensation for non-successful days, which refers to days when the staff fails to complete four surgeries within an eight-hour timeframe. The aim is to allocate this overtime payment towards accommodating a fifth case on successful days. The study based its assumptions on the perspective system discussed in Chapter 5, examining two distinct scenarios. The first scenario is highly optimistic, if the system enables the staff to achieve a 100% SSR. The second scenario is more realistic and achievable, considering a 77% SSR (75th percentile) as the target achievable rate with the system's assistance.

6.2. Methodology

In this chapter, a detailed analysis was conducted to assess the implications of not completing surgeries on time, specifically focusing on the scenario of four arthroplasty surgeries within an 8-hour window, typically scheduled between 7:30 am and 3:30 pm with a 15-minute buffer time, and overtime pay only considered after 3:45 pm. The initial step involved calculating the wasted time and cost associated with unsuccessful days. Subsequently, a linear regression approach was employed to estimate the feasibility of accommodating a fifth case on successful days by leveraging the overtime costs incurred on the unsuccessful days.

This chapter primarily serves as a comprehensive cost-benefit analysis, delineating both the financial implications and the impact on the operating room (OR) throughput that can be realized through the utilization of our previously designed PAS system. It is crucial to note that the focus of this chapter is not on prescribing specific actions to be taken to save time or on detailing the methods to accommodate a fifth case.

Moving forward to the implementation phase in Chapter 7, the practical application of our findings resulted in successfully fitting a fifth case on 50% of the successful days. The accomplishment was made possible through the administration's strategic planning, which involved scheduling 5 cases instead of 4 on deployment days. Essentially, the surgical team was required to handle a fifth case on most days, but it wasn't obligatory to complete it within an 8-hour timeframe. Overtime payment was only necessary if the surgery exceeded the allotted time. This implementation strategy served as a practical demonstration of applying the insights obtained from the cost-benefit analysis. It highlighted the real-world utility of our PAS system in improving operational efficiency within the operating room.

Fundamental research contributions

First demonstration that a 5th surgery is possible if certain AI-driven benchmarks are followed by TOH staff.

6.3. Author Contributions

Author	Contribution to the study
Farid Al Zoubi	Data curation, Investigation, Methodology, Software, Writing-original draft, Validation, Writing-review and editing.
Georges Khalaf	Methodology, Formal analysis, Investigation, Software, Validation, Visualization, Writing-original draft, Writing-review and editing. development and evaluation, co-writing article.
Paul Beaulé	Conceptualization, Methodology, Writing-review and editing
Pascal Fallavollita	Conceptualization, Methodology, Funding acquisition, Project Administration, Resources, Resources, Supervision, Writing-review and editing

6.4. Article

The Following article was accepted and submitted to Frontiers in Digital Health Journal.

frontiers | Frontiers in Digital Health

TYPE Original Research
PUBLISHED 22 September 2023
DOI 10.3389/fdgth.2023.1242214

Check for updates

OPEN ACCESS

EDITED BY
Lisette van Gemert-Pijnen,
University of Twente, Netherlands

REVIEWED BY
Kirti Sundar Sahu,
Canadian Red Cross, Canada
Manisha Mantri,
Center for Development of Advanced Computing (C-DAC), India

*CORRESPONDENCE
Farid Al Zoubi
✉ falzo100@uottawa.ca

RECEIVED 27 June 2023
ACCEPTED 12 September 2023
PUBLISHED 22 September 2023

CITATION
Al Zoubi F, Khalaf G, Beaulé PE and Fallavollita P (2023) Leveraging machine learning and prescriptive analytics to improve operating room throughput.
Front. Digit. Health 5:1242214.
doi: 10.3389/fdgth.2023.1242214

Leveraging machine learning and prescriptive analytics to improve operating room throughput

Farid Al Zoubi[1]*, Georges Khalaf[2], Paul E. Beaulé[3] and Pascal Fallavollita[4]

[1]School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, ON, Canada, [2]The Ottawa-Carleton Institute of Biomedical Engineering (OCIBME), University of Ottawa, Ottawa, ON, Canada, [3]Division of Orthopedic Surgery, Ottawa Hospital Research Institute, Ottawa, ON, Canada, [4]Interdisciplinary School of Health Sciences, University of Ottawa, Ottawa, ON, Canada

Successful days are defined as days when four cases were completed before 3:45pm, and overtime hours are defined as time spent after 3:45pm. Based on these definitions and the 460 unsuccessful days isolated from the dataset, 465 hours, 22 minutes, and 30 seconds total overtime hours were calculated. To reduce the increasing wait lists for hip and knee surgeries, we aim to verify whether it is possible to add a 5th surgery, to the typical 4 arthroplasty surgery per day schedule, without adding extra overtime hours and cost at our clinical institution. To predict 5th cases, 301 successful days were isolated and used to fit linear regression models for each individual day. After using the models' predictions, it was determined that increasing performance to a 77% success rate can lead to approximately 35 extra cases per year, while performing optimally at a 100% success rate can translate to 56 extra cases per year at no extra cost. Overall, this shows the extent of resources wasted by overtime costs, and the potential for their use in reducing long wait times. Future work can explore optimal staffing procedures to account for these extra cases.

KEYWORDS

prescriptive analytics, predictive analytics, machine learning, time forecast, health care efficiency, high volume surgery, operating room throughput

1. Introduction

In Canada, the median wait time for treatment from referral by a general practitioner (GP) was 27.4 weeks (12.6 from GP to specialist, and 14.8 from specialist to treatment) in 2022. This value continues to trend upwards, even relative to pre-determined reasonable wait times (1).When looking at individual specialties, orthopedic surgery not only consistently demonstrates long median wait times over several years, but also has the longest median wait time from specialist to treatment in 2021 (30.2 weeks) and is second only to plastic surgery in 2022 (32.4 vs. 34.3 weeks) (1). Despite having an estimated median reasonable wait time of 15.4 weeks, hip and knee replacement surgeries were still given a Pan-Canadian benchmark wait time of 26 weeks as a maximum, yet it is still lower than the national median wait time of 38.0 weeks in 2022 (1). In Ontario alone, there are currently an estimated 206,000 patients waiting for surgical procedures (2). For orthopedic surgery in Ontario, the median waiting time is 19.9 weeks, 75% greater than the province's reasonable median wait time of 11.4 weeks, leaving an estimated 38,275 patients waiting for orthopedic treatment, 25,372 of which are for arthroplasty surgeries (1).

Before the COVID-19 pandemic, hip and knee replacements were increasing at a rate of 5% per year. During the COVID-19 pandemic, hip and knee replacements between April and

December decreased by 16.1% and 29.8% respectively from 2019 to 2020 due to an abundance of cancellations, creating an excess of waiting patients. In fact, current trends in Canada have led to 138,500 surgeries and estimated inpatient costs of over $1.4 billion a year, imposing a huge burden on the economy, in addition to large backlogs in waiting lists (3). Unfortunately, data indicates that in order to overcome these large backlogs, provinces will need to exceed pre-pandemic rates of surgery, something that has only been accomplished 3 times nationally since the beginning of the pandemic (4).

In Ontario, patients are triaged into different categories based on urgency: priority 4 patients have a target treatment time of 182 days, priority 3 have a target treatment time of 84 days, and priority 2 patients should be treated within 42 days. When evaluating the wait time from the decision for surgery to the surgery itself, only 16% of hip replacement patients and 10% of knee replacement patients are treated within the target time at our institution. For the former treatment, the average wait time for priority 4 patients is 375 days, while that for priority 2 patients is 135 days (not enough data for priority 3), while knee replacement patients yield average wait times of 398 days for priority 4 patients, and 214 days for priority 3 patients (not enough data for priority 2) (5).

Current options that support and enhance the scheduling and efficiency of hospitals come in two main types: clinical and non-clinical methods. The non-clinical methods (6, 7) can be divided into three approaches: firstly, incorporating new resources to enhance the effectiveness of operating rooms (8–11); secondly, utilizing data-driven methods like rhetorical data and descriptive analytics to modify, assess, or reorganize existing hospital resources (12, 13); and thirdly, employing machine learning (ML) solutions specifically designed for operating room optimization. The majority of these solutions have the common goal of foreseeing various aspects of surgical procedures. This involves predicting events both before (14) and after surgery (14–16), as well as accurately estimating the duration of the surgical process itself (17–19).

Expenses can be associated with any of these phases of surgery, which opens up opportunities for cost reduction (20). One way to achieve this is by diminishing the necessary resources through the implementation of self-management application (21). Additionally, certain decision support systems have been employed to simulate various scenarios involving interactions with staff and managers, potentially leading to an increase in the number of cases on certain days (22).

Other Efforts such as 4-joint operating rooms (OR) (i.e., dedicated to serving 4 operations within 8 h) have been implemented to increase throughput and shrink existing waiting lists (23). However, the issue persisted, as the 4-joint room was only able to report a 49% success rate in 2012, indicating a lack of consistent efficiency (23). This is additionally concerning as inefficient use of resources and time contribute to 30% of total healthcare expenditures (14), further emphasizing the need to optimize time and cost. These numbers highlight the burden placed on this hospital, and the pressing need for solutions to reduce the waiting list.

We are of the opinion that we are pioneering the utilization of perspective analytics to compute the expenses linked to overtime pay and forecast whether this sum suffices to accommodate a fifth instance of a 4-joint arthroplasty procedure.

2. Previous work

Our institution has a dedicated orthopedic OR for 4-joint arthroplasty procedures. The OR is specially designed for high-volume arthroplasty surgeries, i.e., partial, and complete joint replacement, and facilitates four procedures each day (from Monday to Friday, excluding Wednesday at which time the ORs start 30 min later for education). Each procedure is subdivided into six stages, including Anesthesia Preparation, patient positioning, surgical procedure, patient exiting the room and turnover, the final stage (see **Figure 1**).

A successful day in this arthroplasty OR is defined as the completion of all four procedures within the allocated time; in this case, the eight hours assigned between 7:30 am and 3:30 pm; however, because there is a 15-min buffer window for overtime pay at our institution, 3:45 pm is used in the proposed methods of this article. The Surgical Success Rate or SSR was the metric designed to keep track of the percentage of successful days in a predetermined period (typically a year).

The original SSR was dismal - 39%, and the overtime cost for our institution was roughly $570,000 annually. Multiple initiatives were introduced to improve this SSR with varying degrees of success (23–25).

Recently, we suggested the most comprehensive solution to this problem—a data-driven, Machine Learning (ML)-based, prescriptive analytics system. It not only predicts the probability of whether a particular day would be successful based on time variables, but also monitors each stage of the procedure in real-time, modifying its prediction if needed, and offers suggestions through a proposed list of actions at a given stage to increase the probability of success (25), as demonstrated in **Figure 2**.

These suggested actions are updated stage-by-stage for each procedure. The multiplicity of suggestions ensures that the surgical team has multiple viable options to choose from, allowing them to leverage their experience and expertise to employ what they believe would be the most positively impactful suggestion in any given circumstance (see **Figure 2**). These real-time proposals allow the system to not just monitor but optimize the procedures and influence the SSR.

The suggestions offered by the ML-based prescriptive analytics system were developed and tested during a comprehensive and highly successful program designed to optimize OR procedures. The program focused on Positive Deviance seminars offered by the most successful surgical professionals in their respective domains. This included surgeons and registered nurses (RNs) with the highest SSR, with the idea of sharing their expertise and best practices with the team to improve overall SSR (8). At our center, during the PD exercise the anesthesiologists refused to participate.

These professionals shared the processes and procedure optimization techniques that allowed them to complete all four surgeries on time (without compromising patient safety) with the

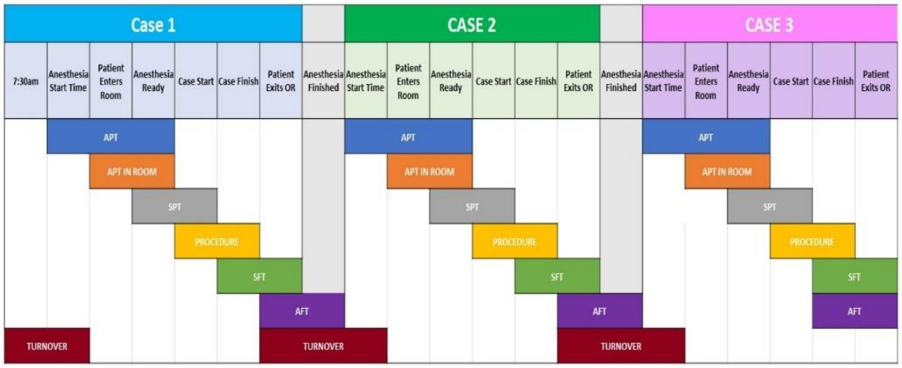


FIGURE 1
An overview of each stage of a procedure, as well as the transition between two (turnover). At the top are the stage markers of a case, while the black region shows the different duration variables used. Acronyms are fully explained in **Table 3**.

FIGURE 2
A visual demonstration of the artificial intelligence model's function. Colors are used to indicate timeliness at each stage of the procedure, while suggestions are provided for ways to maintain or catch up to the desired rate.

rest of the team (see **Table 1**). These processes and techniques were made part of the ML system, which would suggest the right set of actions at various stages of the procedure to optimize the process.

Following the system's suggestions resulted in a significant improvement in individual procedure times and, as a result, an improvement in the overall SSR.

2.1. The benchmarks

The model establishes multiple sets of benchmarks to track success and failure by monitoring the six stages of individual procedure case. More specifically, both stage duration benchmarks and recommendations are produced for any desired SSR. Based on the sets produced, a 77% success rate was defined as the baseline, replacing the default SSR of 39% (25). This SSR was selected as the baseline because its benchmarks were deemed easily achievable by the clinicians and through the leveraging of the prescriptive analytics system (see **Table 2**). In doing so, it leads to improve nearly three out of five (61%) failed days on which the 4 surgeries were not completed on time. These failed days, on average, cost the hospital about 36 min of overtime (more than $2,000 a day).

Another benchmark is the best-case scenario - 100%. This is achievable when the prescriptive analytics system is fully leveraged, and the most potent actions/suggestions are followed to optimize the overall procedure time. This results not only in a successful day, but also with an adequate amount of time left within the eight-hour window. It is this scenario that encouraged us to suggest a follow-up model. Example benchmarks for this scenario as well as others are shown in **Table 2**.

3. Relation to this work

With the development of our ML-based prescriptive analytics system, the ideal scenario would be zero overtime and all

TABLE 1 Examples of suggestions offered by surgeons and nurses at positive deviance seminars to optimize OR procedures.

Surgeons	Nursing
1. Be there from positioning to patient transfer from the table. 2. Have a standardized/protocolized approach for each type of procedure. 3. Anticipate next steps, calling for instruments/implants. 4. Assist with turnover and putting away instruments, but in a way that is supported by nurses. 5. Institute an incentivization for the entire team to be done by 3:30, and that would drive efficiency. 6. Bring the patient into the room for spinal preparation such that instruments may be opened simultaneously (in parallel rather than in series) 7. Anesthesia does the blocks and spinals in the procedure room.	1. Have an engaged, familiar team working together. 2. Have equipment ready to go before patient enters the room. 3. Whole team (nursing, surgery, anesthesia) is present during turnover. 4. Begin putting away instrumentation during closing. 5. Have experienced, knowledgeable scrub nurses who know the steps to the procedure and will know when certain instruments (implants) are needed. 6. Have attendants available to help with turnover. 7. Ensure nurses in the room have received total joint training. 8. Minimize phone call interruptions from pre-op and PACU during the case. 9. Ensure attending available for prep. Make use of free staff in room when prepping/positioning. Ensure no revision of surgical positioning. 10. Need team lead (TL) to have adequate time for training and administration. 11. Ensure improvements in efficiency don't come at cost to patient outcomes.

TABLE 2 Benchmarks established by the AI model for different success rates. The baseline and the optimal scenarios are used for evaluation in this paper.

Scenario	APT (mins)	Case (mins)	AFT (mins)	Turnover (mins)	Success rate
Baseline (75th percentile)	<10.5	<71.5	<20.5	<21.5	77%
Fast procedure	<10.5	53	<20.5	<21.5	93%
	<10.5	64	<20.5	<21.5	89%
Slow procedure	<10.5	>71.5	<20.5	<21.5	59%
Slow turnover	<10.5	<71.5	<20.5	>21.5	69%
Slow anesthesia Preparation	10.5–18.5	<71.6	<20.5	<21.5	64%
Optimal performance	<7	<62.5	≤7	≤20	100%

surgeries completed on time on any given day. However, even if we disregard the anomalies during procedure and preparation, there are several factors preventing this ideal scenario from becoming the norm, including a limited number of high performers and the inevitable concentration of more time-consuming patients on certain days (statistically significant).

This encouraged us to consider the above two benchmarks, 77% and 100% SSR outputs of our perspective system to predict the possibility of completing an additional surgery (i.e., a 5th case) during a successful day. To do this, savings from the decrease in overtime hours and its increased pay will be evaluated for its ability to fund the 5th cases, leading to an increase in throughput with no extra cost. This will ensure fair compensation to the surgical staff for a higher number of overall surgeries because additional surgeries would be covered under the saved overtime that the staff has already been paid for. A flowchart visualizing this process can be found in **Figure 3**.

Even with a rudimentary calculation, the average overtime for four unsuccessful days (36 min times four) will be roughly enough to justify fitting in a fifth surgery on a successful day. With the cost and time justified, the analytics systems will be applied to different time distribution scenarios to identify how many successful five-surgery days are feasible and justifiable. This may require us to consolidate recommendations like scheduling multiple low-risk/less-time-consuming patients on a single day or deferring monitoring in the Post-Anesthesia Care Unit (PACU).

Therefore, the contribution of this article is to implement a predictive method to estimate of likelihood of fitting a 5th=surgery during a successful 4-joint operating room day, using only cost-savings accrued from use of our previous work —an Artificial Intelligence (AI)-based model that produces time-based benchmarks for different success rates. Cost savings from two success rates that will be evaluated: 77%, our realistic baseline, and 100%, optimal performance (25). To do this, we will utilize linear regression by fitting a model to every successful day, generating a histogram of 5th case predictions that will be leveraged to explore different distribution strategies of saved costs.

4. Arthroplasty data set

In order to leverage the existing ML-based prescriptive analytics system for the proposed fifth case simulation, it's imperative to understand the foundational data used to build the original model/system.

4.1. Time period

The data collected for building use in this paper spans from 2012 to 2019. This is enough time to recognize almost all short-term and long-term patterns in data. The statistically significant amount of data naturally lowered variance estimation, leading to more accurate predictions and, consequently, drawing more relevant recommendations.

4.2. Nature of procedures

To streamline the data and identify better optimization techniques, we adhered to non-complex cases and unilateral

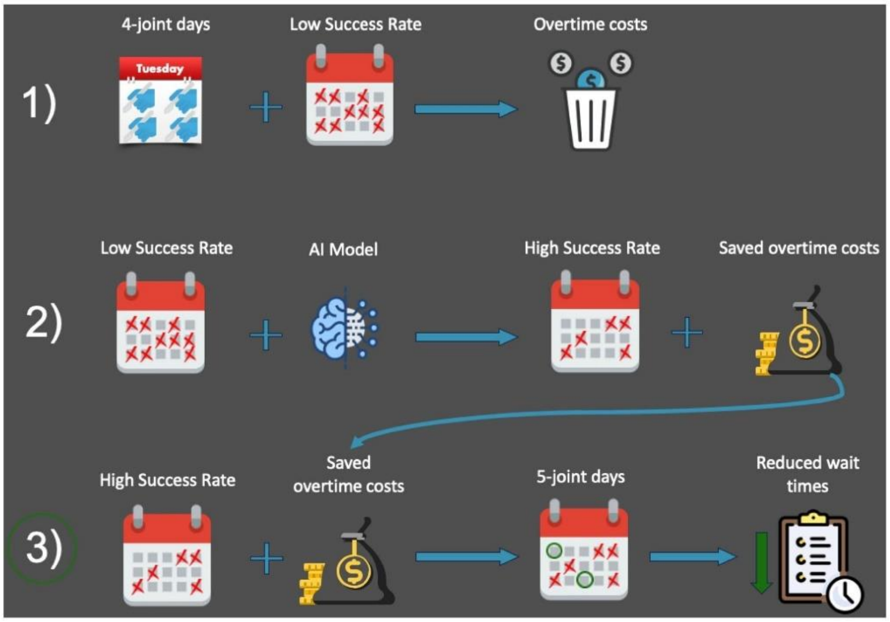


FIGURE 3
A flowchart showing an overview of our research plan. This article focuses on row 3.

surgeries. Our early analyses found that bilateral surgeries take more time, and even if it is not twice as long as unilateral surgeries (such as the actual difference in the surgical procedure), it is statistically significant enough to prevent the on-time completion of four surgeries in a day. It is also important to consider that typically, there is an approximate 4:1 unilateral to bilateral surgery ratio, so the bulk of the procedures were considered. Similarly, complex surgeries where health complications, which may prolong the procedure, are identified beforehand were also excluded from the data set since they are predictable rarities, not the norm.

4.3. Nature of data

The data our machine-learning systems were trained on was both numerical and categorical in nature. The numerical data mostly consisted of timestamps for every stage of the procedure, which were converted to durations to generate a rich number of numerical variables and metrics. The categorical data came from the individuals and the type of surgeries performed. Collectively, the data pool consisting of 40 different variables covering almost all medically relevant details about the patient and procedure, the surgical team performing the procedure, and the necessary time variables; however, the final dataset was filtered down, resulting in 29 of the pool's 40 variables being used (see **Table 3**). Of these 29, Case Number and Out of Room Time will be used in this paper's models, highlighted in **Table 3**.

4.4. Data collection source

The data for most of the identified metrics came from the Surgical Information Management System (SIMS), though some data points came from patient charts and daily notes. This consolidated sourcing of the relevant data prevented the need for integrating different information management systems and overcomplicating the process. This is also one of the factors making this system extrapolatable to different healthcare facilities.

4.5. Treatment of data

Since we already removed anomalies like complicated cases and statistical outliers (bilateral surgeries) that would undermine the pattern recognition and generation of useful insights, the treatment was relatively minimal. The data was cleaned for missing information and incorrect values, both of which represented less than 1% of the observed data set, so the removal was not significant enough to impact the statistical outcome. Regardless of their dissent with median values, rare cases (categorized as non-complex before the procedure began) were kept in the data set. They had a modest impact on the extremes, but not enough to deviate from the trends enough to draw wrong conclusions.

TABLE 3 The selected variables that were present in the dataset used for the development of the AI model.

Time metrics	Staff (team) metrics	Patient metrics	Safety metrics
Anesthesia preparation time (APT)	Surgeon	Campus	90-day readmissions
Anesthesia start time	Anesthesiologist	Type of surgery	Reason for readmission
Time in room	Circulator Nurse 1	Type of anesthesia	Length of stay
Anesthesia ready time	Circulator Nurse 2	Sex	
Anesthesia stop time		Age	
Anesthesia finish time (AFT)		BMI	
Surgical preparation time (SPT)		ASA	
Case start			
Case finish			
Surgery finish time (SFT)			
Turnover			
Surgery (procedure) time			
Time out of room			
Case no			
Date			

Time out of Room and Case No are the features used for this work.

TABLE 4 General patient demographics from the sample dataset.

Number of surgery days (total surgeries)	761 (3,044)
Distribution of male and female patients	1,560 (51.25%) M; 1,484 (48.75%) F
Average patient age	63.2 ± 11.9
Average patient BMI	30 ± 5

4.6. Demographics of patients and other quantifiable

Table 4 contains required information about demographics of patients and some other quantifiable.

4.7. Descriptive analytics

Of the 761 4-joint operation days, 301 were successful (4 operations before 15:45), marking a 39.55% success rate. In these successful days, there was a total 97 h and 49 min of spare time (time between the end of the final case and 15:45), averaging to 19 min and 30 s per day. Overtime-cost hours were calculated using the remaining 460 unsuccessful days by multiplying the number of overtime hours (hours worked past 15:45) by 1.5 (the paid overtime rate). Doing so reveals a total 465 h, 22 min, and 30 s overtime-cost hours, leading to an average of one hour per unsuccessful day.

Each case's out of room time, which is the time at which the patient is taken out of the operating room, is plotted on a histogram, showing distributions of successful cases and all cases (see Figure 4). Peaks on the graph belong to individual distributions of one of the four cases in a day. The distribution of all 4th cases shows a tail that extends past 8 pm, marking over 4 h of overtime on some days. In general, the skew of all cases appears to increase to the right with each subsequent case number, while that of successful cases appear to do so minimally to the left. In line with the trend of increasing skews, the spread of each distribution also increases with each subsequent case number, going from having a standard deviation of 00:17 m:44 s for first cases to 00:42 m:05 s for fourth cases, as shown in

FIGURE 4
A histogram of out of room times showing all individual cases, and successful cases. The x-axis shows bins of times at which patients have been escorted out of the operation room, while the y-axis shows the number of cases that are in each bin.

TABLE 5 The standard deviations of successful cases and all cases.

	All cases	Successful cases
1st case	00:17:44	00:12:07
2nd case	00:26:14	00:17:01
3rd case	00:31:26	00:18:56
4th case	00:42:05	00:16:51*

*Reduced by a cut of the distribution due to the limit at 15:45.

Table 5. Table 5 also shows a stark difference between successful cases and all cases, as the former is much more consistent with their times, as demonstrated by their smaller standard deviations.

Individual successful days were isolated, and their out of room times were each graphed against their case number. Figure 5 shows an example of 4 different days, all of which demonstrate the linear nature of out of room times for a specific day. This observed trend inspired our method of predicting 5th cases as described in the Methodology.

5. Methodology

The data was first divided into successful and unsuccessful groups: the successful group contains all days where the fourth case was completed (as defined by its out of room time) at or before 15:45, while the unsuccessful group contained all days where the 4th case was completed after 15:45.

As previously mentioned, there is an average of 19 min and 30 s of spare time per successful day; to ensure that this spare time is used and that staff do not end their days later than needed, it was decided that 5th cases will only be added to successful days. This decision is further reinforced by the successful group's more consistent trends, which facilitates model training to yield more accurate results. Thus, the successful group is used for the prediction of 5th cases and their potential addition, while the unsuccessful group is used for the calculation of overtime-cost hours and the distribution of their hypothetical savings among all days.

To predict 5th case out of room times, the linear nature of each successful day's cases is leveraged (see Figure 5). All 301 successful days are isolated, along with their 4 cases. For each isolated day, a linear regression model is used to fit the out of room times with case number as the independent variable, and a 5th case prediction is generated from each one. Following the prediction for each day, a distribution of 5th case predictions is produced (shown in Figure 6). This distribution will be used to evaluate the potential of adding extra cases by using previous cost savings calculated from unsuccessful days. The use of linear regression is advantageous due to its interpretability, simplicity, and its ability to make predictions without ground truth data. Due to the nature of the problem, no 5th case ground truth data is available, limiting the scope of methods to choose from. Linear regression overcomes this problem by not requiring the desired input to be in the training set. Given the approach of isolating successful days, each day's model can be individually analyzed and adjusted, with its slope representing that day's average case duration. Further, linear regression accounts for the day's start time by accounting for the first case, and naturally produces variability in case durations without compromising overall accuracy by fitting to the existing variability within the dataset. Finally, this method allows for the prediction of further cases (ex. 6th, 7th, etc.) if needed by simply changing the input variable. The linear regression equation is shown below, where w_1 denote the trainable weights, x is the inputted case number, and y is the associated Out of Room time. Practically, w_1 reveals a given day's

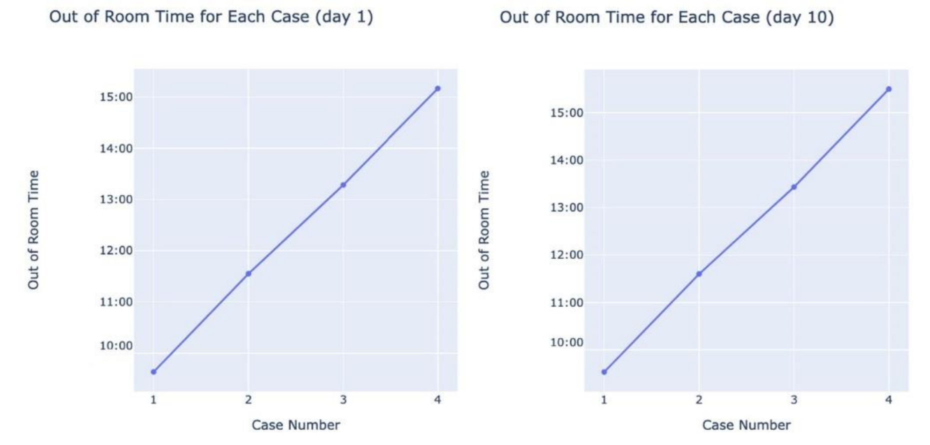


FIGURE 5
Random examples of successful days' out of room times with case numbers on x and out of room times on y.

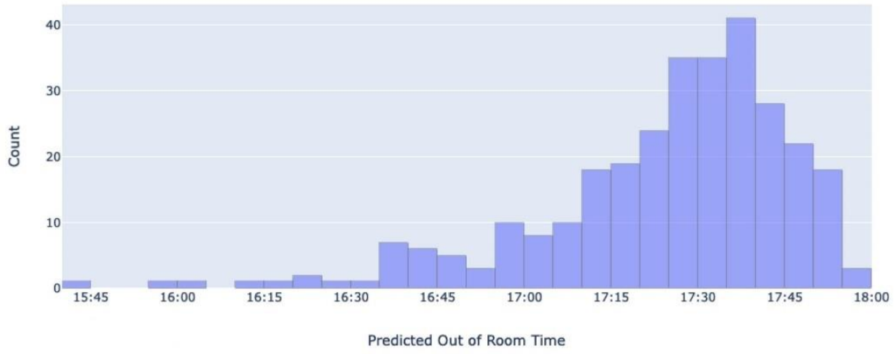


FIGURE 6
A histogram of all predicted 5th case out of room times. The x-axis represents different bins of predicted 5th case out of room times (time at which patients are escorted out of the operation room), while the y-axis shows the number of predicted cases that are in each bin.

average case duration, while w_0 accounts for varying start times between days.

$$y = w_1 x + w_0 \tag{1}$$

Calculations for overtime-cost hours saved (OCHS) were made using this equation:

$$\text{OCHS} = 465.375 - 761x(1 - y) \tag{2}$$

Where x is the previously calculated overtime-cost hours per unsuccessful day, y is the new success rate, 465.375 represents the current number of overtime-cost hours (based on the 39.55% success rate), and 761 is the total number of days in the dataset. Thus, a 100% success rate would lead to saving 465.375 overtime-cost hours. The calculated OCHS is then divided in 3 ways: daily (761 days), bi-daily, and once a week (4-day work weeks), each producing a different 5th case success time benchmark. Once done, the predicted 5th cases are used to generate success rates for each success time and calculate the number of potential extra cases delivered at no extra cost.

6. Validation

Because there is no ground truth data, two methods of validation are employed. Both methods rely on Mean Absolute Error (MAE), which is described by the following equation:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{3}$$

Where n is the sample size, y is the actual value, and $\hat{y}$ is the predicted value. MAE is used because of its interpretability, as it provides the actual mean time difference between the generated values and the ground truth.

The first method is the prediction of 3rd and 4th case out of room times, so that MAEs can be generated from their existing ground truth data. To do this, the same procedure that is used to predict 5th cases is also used to predict 3rd and 4th, with the difference being that the lines are only fit to the first 2 and 3 cases of each successful day respectively. Once all the errors are calculated, histograms and 95% mean confidence intervals are produced for the MAEs to gain a deeper understanding of the model's performance. Through this method, we are leveraging existing ground truth within the data to produce MAEs that can be used to infer that of the prediction of 5th cases.

The second method is used to give an impression of how well the lines fit to the existing data. Each day's model has their MAE calculated using the points on the line and the actual 4 case out of room times. From there, another distribution and 95% mean confidence interval is generated for further insight into linear regression's performance on the dataset, as well as the 5th case predictions' errors themselves.

7. Results

All predictions were compiled and visualized as a distribution (**Figure 6**). The predictions have a mean time of 17:24:17 (95% CI = 17:21:46, 17:26:48) a median of 17:29:30, and a standard deviation of 22 min and 10 s. Of the 301 predictions, 256 of them (85.7%) fall below the 2-h mark (17:45), while all 301 (100%) are predicted to end before 18:00, as the latest predicted time is 17:56:30.

After training 3rd and 4th case-predicting models, mean absolute error values of 13 m:40 s and 14 m:13 s minutes respectively were calculated. A distribution of every day's error for each model is shown in **Figure 7**. These distributions yield 95% mean confidence intervals of 00 h:12 m:28 s, 00 h:14 m:50 s and 00 h:12 m:53 s, 00 h:15 m:31 s for the 3rd and 4th case models respectively.

Similar outputs were produced for the second method; the mean of all daily MAEs was 4:45 min, with a 95% mean confidence interval of 00 h:04 m:25 s, 00 h:05 m:05 s and a standard deviation of 2:55 min. The histogram of all daily MAEs can be found at **Figure 8**.

8. Linear regression compared to other techniques

While evaluating different approaches, the characteristics of our dataset significantly limited the choices available to us. Firstly, the absence of ground truth data for the fifth case eliminated the option of employing machine learning techniques reliant on labeled training data. Secondly, we identified a pronounced linear pattern in successful outcomes, prompting the adoption of linear regression. Although we explored alternative regression methodologies, it became evident that none were viable for our particular scenario where each day required a distinct model. Bayesian regression, for instance, necessitated data distributions to derive information, but this method proved unsuitable given the small dataset of four data points per model across each day.

Given the solvability of linear regression, we disregarded models employing gradient descent or regularization techniques as unnecessary. Nonetheless, for the purpose of comparative analysis, we contemplated incorporating the average duration of each case (119 min). To achieve this, we added the time taken for the fourth case from the available room time and assigned this duration to the fifth case, assuming a success time of

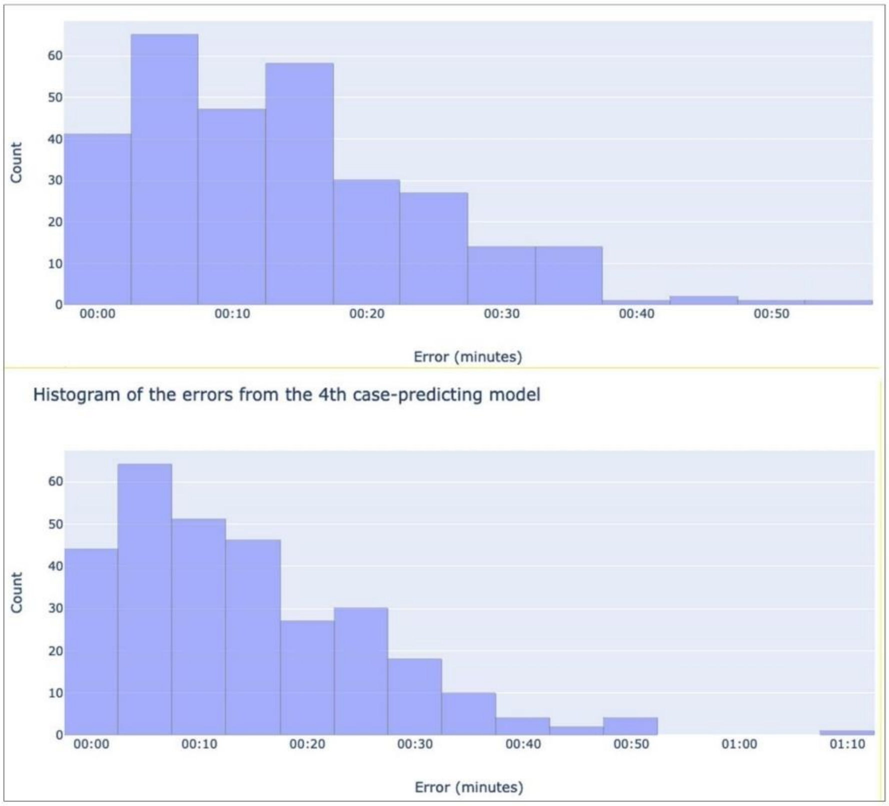


FIGURE 7
Histograms showing the distribution of the absolute values of each recorded error for the 3rd case-predicting model (top), and the 4th-case predicting model (bottom). The x-axes denote bins of absolute error values in minutes, while the y-axes show the number of predictions in each bin.

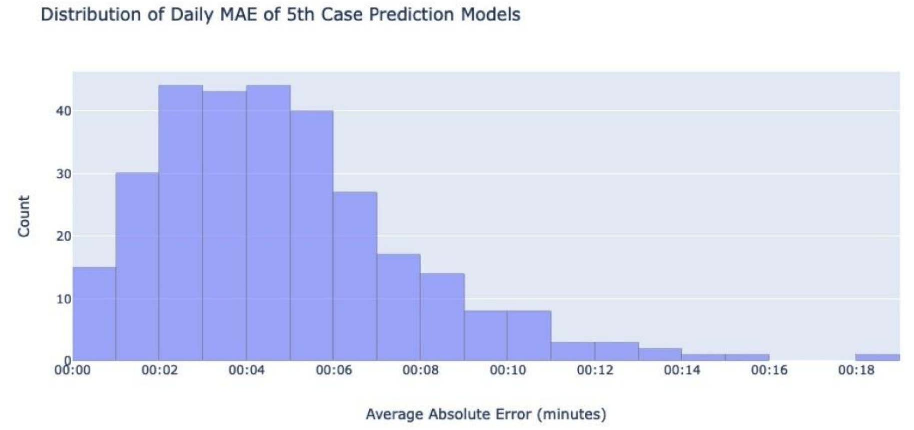


FIGURE 8
A histogram showing a distribution of the absolute values of each day's MAE. MAEs were calculated using the differences of each ground-truth case and their fitted lines. The x-axis shows bins of mean absolute errors for any given day's 5th case-predicting model, while the y-axis counts the number of days in each bin.

5:45 pm. This hypothetical adjustment yielded a 100% success rate, as all fourth cases concluded prior to 3:45 pm. It is important to note that this outcome is unrealistic and contradicts our understanding of the situation. Furthermore, the identical distribution between the fourth and fifth cases is also unrealistic, particularly when considering that the standard deviation increases with each successive case number.

9. Predicting the potential to Fit a 5th case during successful surgery days

The predictions were considered under two potential scenarios: 77% success rate, and 100% success rate. Using Formula 2, achieving a 77% success rate would yield hypothetical savings of 288 h:17 m:50 s overtime-cost hours, which is approximately 38 h and 26 min per year. Distributing these hours daily leads to 22 m:44 s extra minutes per 5th case day, which when added to the original end time of 15:45, would produce a new end time of 16h:07 m:44 s. Based on predictions, 5th cases would be completed at a 1.00% success rate for that time. This extra time is doubled when distributed bi-daily to 45:28 min per day, marking a surgery end time of 16:30:28 and more than doubling the 5th case success rate to 2.66%. Finally, given that 4-joint days are only run 4 days a week, the extra time is once again doubled when pooling them for a weekly 5th case. Doing so yields 90:56 extra minutes per 5th case day, for a surgery end time of 17:15:56 and predicted success rate of 26.25%. However, because an end time of 18:00 (135 extra minute) has a predicted success rate of 100%, one week can be skipped to split its extra time among the following two weeks and lead to two days with a predicted 100% 5th case success rate every 3 weeks. Ultimately, this sums up to approximately 35 extra cases per year at no extra cost. A map of potential distributions at a 77% success rate is shown in **Figure 9**.

Under the situation where a 100% success rate is achieved, all 465 h:22 m:30 s overtime-cost hours would be saved. With a daily split, this amounts to 36 m:42 s per 5th case day, an end time of 16 h:21 m:42 s, and a predicted 5th case success rate of 1.99%. Distributed bi-daily, these values increase to 73 m:21 s, 16:58:24, and 12.6% respectively. When divided weekly, 146:48 extra minutes, adding to an end time of 18:11:48, are given per day, producing a predicted success rate of 100% with a minimum of 10 min to spare. These spared minutes can be pooled to contribute to another 4 cases per year, leading to a total of 56 potential cases per year (assuming the hospital runs all year long). A schematic showing these results is found at **Figure 10**.

10. Discussion

Considering poor 4-joint day success rates, our previous work sought the development of an AI model that provides benchmarks to achieve a certain success rate. Given our institutions' current wait list issues with hip and knee replacement surgeries, 5th cases were predicted to evaluate the potential of their addition using only overtime savings from an increase in success rate. This can lead not only to better staffing efficiency, but higher surgical throughput to help reduce the waiting list as well.

After completing 5th case out of room time predictions, a mean out of room time of 17:24:17 and a median of 17:29:30 were predicted. A negatively skewed distribution was expected due to

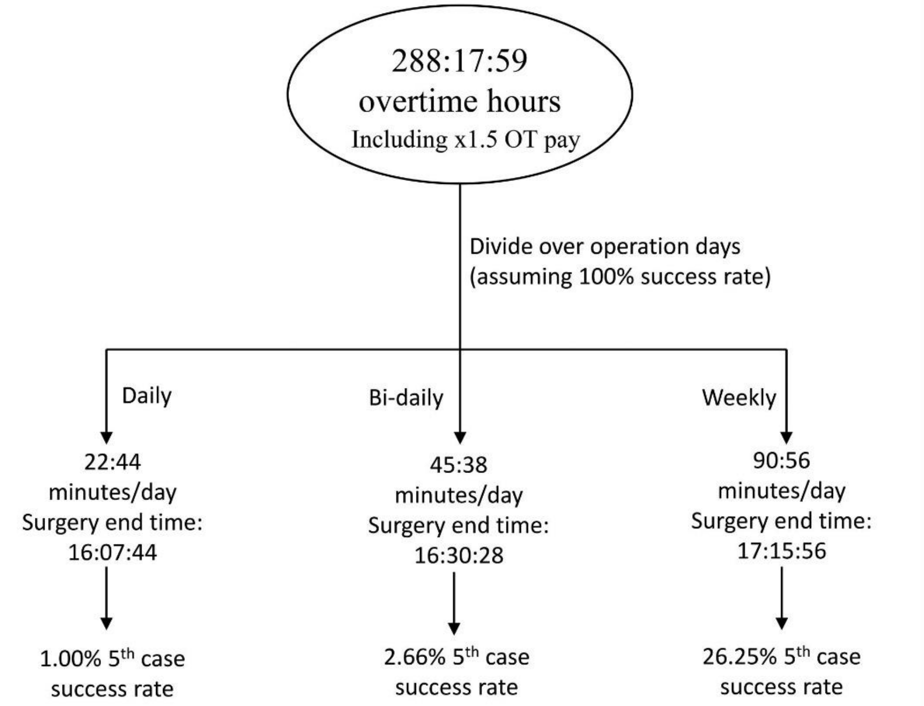


FIGURE 9
A schematic demonstrating the division of saved overtime-cost hours from performance at a 77% success rate. There is only a 1% chance of fitting a 5th case successfully each day, and up to 26.25% chance to add a 5th case per week.

the same trend with 4th case out of room times. The predictions' standard deviation of 22 min and 10 s also falls within expectation, as it increased relative to previous ground truth cases, as shown in Table 5. Despite the average successful case time of an hour and 59 min, not all predicted 5th cases (85.7%) fall below the 2-h mark (17:45) while all 100% are predicted to end before 18:00, as the latest predicted time is 17:56:30. These findings highlight the presence of nuances in trying to find the optimal balance between time added and prevention of further overtime waste.

When applying the predictions to evaluate how cost-savings can be used to fund 5th cases, two contexts are considered: performance at a 77% success rate, our baseline rate that is deemed achievable and realistic by the clinicians (based on the model's benchmarks), and performance at a 100% success rate, which is the ideal, best-case scenario. In both cases, it seemed preferable to pool the hours at different intervals in order to maximize throughput while minimizing the risk of overtime costs.

At a 77% success rate, the time saved would lead to sub-3% success rates when divided daily or bi-daily. Although this would lead to much higher throughput, it would do so at the cost of many overtime hours, where salary is increased by a factor of 1.5. This case is also true when dividing the hours weekly, as 90 extra minutes leads to a 5th case success rate of only 26.25%. Intuitively, this gives the impression that extra time can be pooled further to guarantee one extra case per month with no extra cost; however, simply adding another 45 min is enough to guarantee success based on the predictions, something that can be done for two weeks by skipping one. In other words, when distributing hours to one day a week, skipping one week leads to a 100% 5th success rate in the following two. Overall, this means that approximately 35 extra cases per year at our institution can be funded solely from the savings accrued by increasing performance to a 77% success rate.

When performing optimally (100%), results were similar, as pooling the saved costs also drastically reduced 5th case overtime costs. Daily and bi-daily distribution of saved costs yielded 5th case success rates of 1.99% and 12.6% respectively, while weekly pooling of saved costs allowed for 100% success rate, with a minimum 10 min to spare. In total, savings can be optimized to project 56 extra cases per year, meaning that 56 cases worth of overtime-cost hours are currently being spent due to inefficient performance at our institution (39.55% success rate).

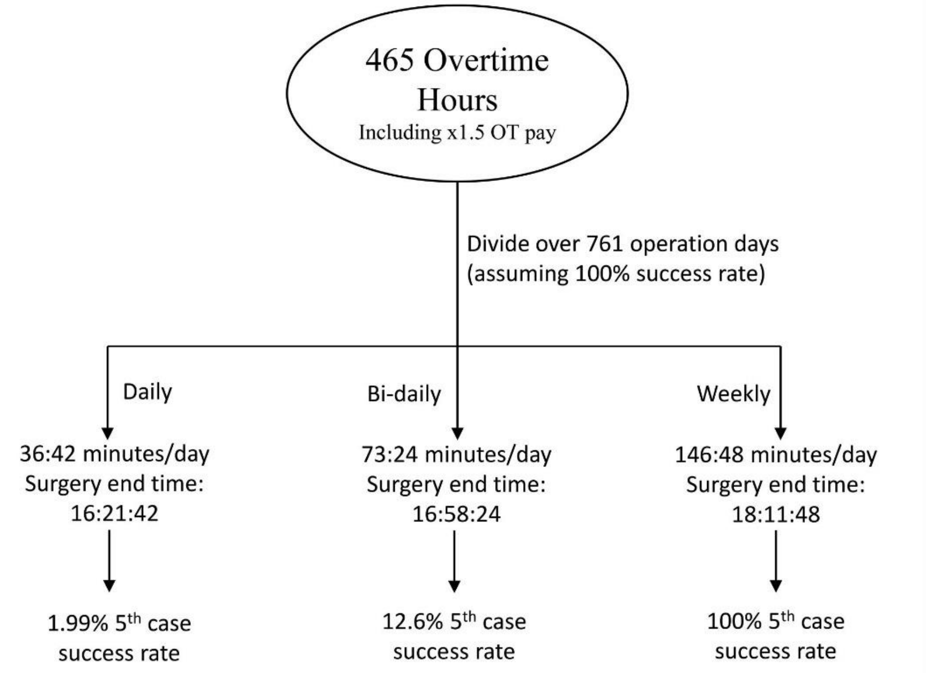


FIGURE 10
A schematic demonstrating the division of saved overtime-cost hours from performance at a 100% success rate.

However, one limitation with our output is the inability to conclusively measure prediction accuracy due to the lack of ground truth 5th case data. Instead, existing data was leveraged to infer the models' accuracy. The first method of doing so, training models to predict 3rd and 4th cases, yielded mean absolute error values of 13 m:40 s and 14 m:13 s minutes respectively. Based on the histograms, most errors are smaller than the means (see **Figure 7**), with a few large outliers. This is deemed acceptable as these errors can represent variations in case durations that exist in the dataset, making a more representative distribution of predictions. The second method looked instead at how well linear regression fit to the existing data, yielding a mean absolute error value of 4 m:45 s. As with the previous method, the distribution of errors shows a positive skew, indicating that most of the errors are below the mean with a few large outliers. Overall, linear regression fit well to the trends of the dataset. However, as mentioned, the lack of any ground truth 5th case data makes this evaluation inferential, as a more direct evaluation cannot be made.

Another limitation is the uncertainty of how the data might change once the AI model (9) is implemented. Whether use of the model would work by improving the speed of all cases, reducing the number of slow cases, or simply streamline case durations so that they are more consistent is unknown, and could impact the distribution of predictions. Fortunately, it is likely that the use of the model would shift the distribution to the left, potentially making the current evaluation a pessimistic one.

Despite these limitations, we propose a simple, effective, and reproducible method of calculating potential throughput gains with no extra cost as a result of improved performance efficiency. In our case, this improvement relies on the success of a benchmark-establishing AI model developed by members of our team. Furthermore, the gains are only attainable with the modification of staffing procedures so that longer days are had without spending overtime rates; one example for this is to benefit from staff that show up late, and who can stay late, by having them stay longer for the fifth case. This work also opens many avenues of future research: reproduction of this work after implementation of the AI model may produce further refinements to cost-free throughput enhancement depending on how the model affects all cases, and whether it improves successful case durations as well. Research into how staffing can best be modified to account for extra cases could also offer another level of optimization, and a potential area of healthcare reform.

11. Conclusion

Due to the COVID-19 pandemic, the Canadian healthcare system was burdened with long hip and knee replacement wait lists and extra costs as a result of cancelled procedures. We aimed to leverage the savings that would be accrued from the use of our AI model to increase surgical throughput with no extra costs. To do this, linear regression models were used to predict 5th case out of room times that served as benchmarks to estimate success rates at different 5th case success times. Success times were determined by distributions of hypothetical overtime-cost savings that would be accrued using the AI model. Previously, our institution operated at a 39.55% success rate. Overall, it was found that increasing to a 77% rate can lead to approx. 35 extra cases per year funded solely by the savings acquired, while operating at a 100% success rate can lead to 56 additional cases per year. Future work can look at the optimization of staffing procedures to account for extra hours with no overtime pay, and the effects of the AI model on all case durations.

Data availability statement

The datasets presented in this article are not readily available because Data has not gone through institutional hurdles to make it open source and available for all. Requests to access the datasets should be directed to FA, falzo100@uottawa.ca.

Ethics statement

The project falls within the context of quality initiative, quality improvement, quality assurance, and/or program evaluation. Consequently, the ethics review was not required at our institution.

Author contributions

FA: Data curation, Investigation, Methodology, Software, Writing—original draft, Validation, Writing—review and editing. GK: Methodology, Formal analysis, Investigation, Software, Validation, Visualization, Writing—original draft, Writing—review and editing. PB: Conceptualization, Methodology, Writing—review and editing. PF: Conceptualization, Methodology, Funding acquisition, Project Administration, Resources, Resources, Supervision, Writing—review and editing. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Barua B, Rovere MC, Skinner BJ. Waiting your turn: wait times for health care in Canada 2010 report. *SSRN Electronic J.* (2012):7–8. doi: 10.2139/ssrn.1783079

2. Wait Time Alliance. *Time to close the gap: report card on wait times in Canada*. Available at: https://www.cag-acg.org/images/quality/wta_report_card2014.pdf (Accessed June, 2014).

3. Canadian Institute for Health Information. Hip and Knee Replacements in Canada: CJRR Annual Statistics Summary, 2018–2019. *Canadian Joint Replacement Registry (CJRR) Annual Report*. Ottawa, ON: CIHI (2020). Available at: https://secure.cihi.ca/free_products/CJRR-annual-statistics-hip-knee-2018-2019-report-en.pdf

4. Canadian Institute for Health Information. *Surgeries impacted by COVID-19: An update on volumes and wait times*. (2023). Available at: https://www.cihi.ca/en/surgeries-impacted-by-covid-19-an-update-on-volumes-and-wait-times

5. Ontario Health. *Wait Time Results for Surgery*. Ottawa (2023). Available at: https://www.ontariohealth.ca/public-reporting/wait-times

6. 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) 22(1):151–66. doi: 10.1186/s12911-022-01893-8

7. Devi SP, Rao KS, Sangeetha SS. Prediction of surgery times and scheduling of operation theaters in optholmology department. *J Med Syst*. (2012) 36(2):415–30. doi: 10.1007/s10916-010-9486-z

8. Yeoh C, Mascarenhas J, Tan KS, Tollinche L. Real-time locating systems and the effects on efficiency of anesthesiologists. *J Clin Anesth Pain Manag*. (2018) 2(1):165–72. doi: 10.1057/hs.2014.6

9. Bender JS, Nicolescu TO, Hollingsworth SB, Murer K, Wallace KR, Ertl WJ. Improving operating room efficiency via an interprofessional approach. *Am J Surg*. (2015) 209(3):447–50. doi: 10.1016/j.amjsurg.2014.12.007

10. Lo EY, Bowler J, Lines T, Melton C, Volkmer R, Majekodunmi T, et al. Operating room efficiency and cost reduction in shoulder arthroplasty: is there an advantage of a dedicated operating room team? *Seminars Arthroplasty: JSES*. (2021) 31(1):125–30. doi: 10.1053/j.sart.2020.11.002

11. Crawford D, Lombardi A, Berend K. Improving operating room efficiency with single-use disposable instruments for total knee arthroplasty. *Surg Technol Online*. (2022) 40:353–6. doi: 10.52198/22.STI.40.OS1553

12. Porta CR, Foster A, Causey MW, Cordier P, Ozbirn R, Bolt S, Allison D, et al. Operating room efficiency improvement after implementation of a postoperative team assessment. *J Surg Res*. (2013) 180(1):15–20. doi: 10.1016/j.jss.2012.12.004

13. Dyas AR, Lovell KM, Balentine CJ, Wang TN, Porterfield JR Jr, Chen H, et al. Reducing cost and improving operating room efficiency: examination of surgical instrument processing. *J Surg Res*. (2018) 229:15–9. doi: 10.1016/j.jss.2018.03.038

14. Fairley M, Scheinker D, Brandeau ML. Improving the efficiency of the operating room environment with an optimization and machine learning model. *Health Care Manag Sci*. (2019) 22(4):756–67. doi: 10.1007/s10729-018-9457-3

15. Maheshwari K, You J, Cummings KC 3rd, Argalious M, Sessler DI, Kurz A, et al. Attempted development of a tool to predict anesthesia preparation time from patient-related and procedure-related characteristics. *Anesth Analg*. (2017) 125(2):580–92. doi: 10.1213/ANE.0000000000002018

16. Schiele J, Koperna T, Brunner JO. Predicting intensive care unit bed occupancy for integrated operating room scheduling via neural networks. *Nav Res Logist*. (2021) 68(1):65–88. doi: 10.1002/nav.21929

17. Bartek MA, Saxena RC, Solomon S, Fong CT, Behara LD, Venigandla R, et al. Improving operating room efficiency: machine learning approach to predict case-time duration. *J Am Coll Surg*. (2019) 229(4):346–54.e3. doi: 10.1016/j.jamcollsurg.2019.05.029

18. Strömblad CT, Baxter-King RG, Meisami A, Yee S-J, Levine MR, Ostrovsky A, et al. Effect of a predictive model on planned surgical duration accuracy, patient wait time, and use of presurgical resources. *JAMA Surg*. (2021) 156(4):315. doi: 10.1001/jamasurg.2020.6361

19. Rozario N, Rozario D. Can machine learning optimize the efficiency of the operating room in the era of COVID-19? *Can J Surg*. (2020) 63(6):E527–9. doi: 10.1503/cjs.016520

20. Christou CD, Athanasiadou EC, Tooulias AI, Tzamalis A, Tsoulfas G. The process of estimating the cost of surgery: providing a practical framework for surgeons. *Int J Health Plan Manag*. (2022) 37(4):1926–40. doi: 10.1002/hpm.3431

21. Gollish JD, Pereira L, MacLeod AM, Wainwright A, Kennedy D, Robarts S, et al. Myhip&knee: improving patient engagement and self-management through Mobile technology. *Healthc Q*. (2019) 22(2):63–7. doi: 10.12927/hcq.2019.25902

22. Persson M, Hvitfeldt-Forsberg H, Unbeck M, Sköldenberg OG, Stark A, Pettersson PK, et al. Operational strategies to manage non-elective orthopaedic surgical flows: a simulation modelling study. *BMJ Open*. (2017) 7(4):1–9. doi: 10.1136/bmjopen-2016-013303

23. Beaulé PE, Frombach AA, Ryu J-J. Working toward benchmarks in orthopedic OR efficiency for joint replacement surgery in an academic centre. *Can J Surg*. (2015) 58(6):408–13. doi: 10.1503/cjs.001215

24. Gold R. Use of multidisciplinary positive deviance seminars to improve efficiency in a high-volume arthroplasty practice; a pilot study. *Can J Surg*. (2022) 66:1–7. doi: 10.1503/cjs.018121.

25. Al Zoubi F, Gold R, Poitras S, Kreviazuk C, Brillinger J, Fallavollita P, et al. Artificial intelligence-driven prescriptive model to optimize team efficiency in a high-volume primary arthroplasty practice. *Int Orthop*. (2022) 47:343–50. doi: 10.1007/s00264-022-05475-1

Chapter 7: Validation of the Prescriptive Analytics System in clinical practice

7.1. Summary

In this chapter of my thesis, the central focus revolves around the comparison between theoretical concepts and practical outcomes. I assess the performance of my algorithms and evaluate the efficacy of implementing the WTMF into an integrated PAS, a prototype. Additionally, I validate the feasibility of accommodating an additional case in real-world surgical practices in Appendix D.

7.2. Methodology

The primary objective of this thesis is to formulate and implement a machine learning solution aimed at enhancing the operating room throughput for arthroplasty procedures by optimizing overall efficiency within the surgical team. Among the three frameworks proposed in Chapter 3, specifically designed to enhance operating room throughput through different methods, the WTMF stands out and has been selected for further development and advancement in the implementation phase.

The implementation process focused on creating a prototype of the PAS system, requiring only the essential components of the whole components outlined in Appendix B to validate the concept. These include the machine learning output of benchmarks and the estimated SSR, recommendations necessary to achieve specific SSR scenarios elucidated in Chapter 6, and a rudimentary monitoring system.

Surgical team selection was random, with volunteers participating, some having attended Positive Deviance (PD) seminars, while others had not. Varied records existed for different team members, and awareness of the project varied among participants. The selection bias analysis is in Appendix C.

The selection of operating days, occurring exclusively on Saturdays, was not within our control. This decision aligned with the Ontario government's initiative to augment surgical capacity, with TOH dedicating two operating rooms specifically for hip and knee replacement surgeries on Saturdays. Despite this constraint, the implementation aimed to showcase the practicality and effectiveness of the PAS system under real-world conditions.

7.3. Fundamental research contributions

Deployment of Artificial Intelligence PAS in Orthopedic Surgery.

7.4. Author Contributions

The tangible impact will only be evident upon the publication of the paper. The journal provides designated categories for author contributions during submission. The table we supplied as our contribution is outlined below.

Author	**Contribution to the study**
Farid Al Zoubi	Data curation, Investigation, Methodology, Software, Writing - original draft, Visualization, Validation, Writing.
Koorosh Kashanian	Data Collection, proofreading
Paul Beaulé	Conceptualization, proofreading
Pascal Fallavollita	Conceptualization, Project Administration, Resources, Supervision, Writing – review and editing.

7.5. Article

The Following article was accepted to Frontiers in Artificial Intelligence Medicine and Public Health.

frontiers | Frontiers in Artificial Intelligence

TYPE Community Case Study
PUBLISHED 31 January 2024
DOI 10.3389/frai.2024.1342234

First deployment of artificial intelligence recommendations in orthopedic surgery

Farid Al Zoubi[1]*, Koorosh Kashanian[2], Paul Beaule[2] and Pascal Fallavollita[3]

[1] School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, ON, Canada,
[2] Division of Orthopedic Surgery, Ottawa Hospital Research Institute, Ottawa, ON, Canada,
[3] Interdisciplinary School of Health Sciences, University of Ottawa, Ottawa, ON, Canada

Check for updates

OPEN ACCESS

EDITED BY
Lirong Wang,
University of Pittsburgh, United States

REVIEWED BY
Liangyuan Na,
Massachusetts Institute of Technology, United States
Haohan Wang,
University of Illinois at Urbana-Champaign, United States

*CORRESPONDENCE
Farid Al Zoubi
✉ falzo100@uottawa.ca

RECEIVED 21 November 2023
ACCEPTED 17 January 2024
PUBLISHED 31 January 2024

CITATION
Al Zoubi F, Kashanian K, Beaule P and Fallavollita P (2024) First deployment of artificial intelligence recommendations in orthopedic surgery.
Front. Artif. Intell. 7:1342234.
doi: 10.3389/frai.2024.1342234

Scant research has delved into the non-clinical facets of artificial intelligence (AI), concentrating on leveraging data to enhance the efficiency of healthcare systems and operating rooms. Notably, there is a gap in the literature regarding the implementation and outcomes of AI solutions. The absence of published results demonstrating the practical application and effectiveness of AI in domains beyond clinical settings, particularly in the field of surgery, served as the impetus for our undertaking in this area. Within the realm of non-clinical strategies aimed at enhancing operating room efficiency, we characterize OR efficiency as the capacity to successfully perform four uncomplicated arthroplasty surgeries within an 8-h timeframe. This Community Case Study addresses this gap by presenting the results of incorporating AI recommendations at our clinical institute on 228 patient arthroplasty surgeries. The implementation of a prescriptive analytics system (PAS), utilizing supervised machine learning techniques, led to a significant improvement in the overall efficiency of the operating room, increasing it from 39 to 93%. This noteworthy achievement highlights the impact of AI in optimizing surgery workflows.

KEYWORDS

machine learning, healthcare system, operating room efficiency, orthopedic surgery, prescriptive analytics

1 Introduction

The application of artificial intelligence (AI) in healthcare has seen remarkable advancements since its inception, encompassing various areas such as diagnosis, genetics, prognosis, and drug discovery (Bohr and Memarzadeh, 2020). Despite the vast potential, the predominant focus has often been on the clinical aspect of AI, specifically targeting patient-centered applications rather than addressing the broader spectrum of healthcare processes (Maier-Hein et al., 2022).

A noticeable trend in recent developments revolves around the competitive drive to enhance surgical procedures through the integration of AI. Robotic technologies, in particular, have garnered significant attention among researchers and AI inventors, reflecting a concentrated effort to revolutionize surgical practices (Nwoye et al., 2023).

In contrast, a fewer body of work has delved into the non-clinical facets of AI, concentrating on leveraging data to enhance the overall healthcare system and operating room efficiency.

This includes initiatives aimed at improving team efficiency (Al Zoubi et al., 2022), optimizing patient appointment scheduling (Erekat et al., 2020), and predicting overall procedural durations either partially (Bartek et al., 2019; Schiele et al., 2021; Strömblad et al., 2021) or comprehensively (Al Zoubi et al., 2023).

Despite these advancements, a notable gap exists in the literature pertaining to the deployment and outcomes of AI solutions (Jiang et al., 2021). As of now, there is a lack of published results showcasing the practical implementation and effectiveness of AI solutions in areas beyond the clinical realm, particularly in the field of surgery.

Some studies that have applied AI in healthcare have examined the obstacles that may hinder progress in this field. These challenges might stem from the interdisciplinary nature of AI solutions in healthcare (Safavi et al., 2019), a limited availability of interpretable machine learning models, resistance from medical unions and associations, extended processing times (Bertsimas et al., 2022), and the complexities associated with decision-making and bureaucratic processes involving multiple stakeholders (Hu et al., 2021).

Within the realm of non-clinical strategies aimed at enhancing operating room efficiency, we characterize OR efficiency as the capacity to successfully perform four uncomplicated arthroplasty surgeries within an 8-h timeframe. In this Community Case Study, we disclose the outcomes arising from the incorporation of AI recommendations at The Ottawa Hospital (TOH), Ottawa, Canada leading to a substantial enhancement in the overall efficiency of the arthroplasty operating room, elevating it from 39 to 93%. This remarkable accomplishment was realized by deploying a prescriptive analytics system (PAS) that utilizes supervised machine learning techniques to generate benchmarks specifically tailored for arthroplasty surgery workflows.

2 Methodology

2.1 Problem statement

Our institution has a specialized orthopedic operating room dedicated to 4-joint arthroplasty procedures. This facility is designed specifically for high-volume arthroplasty surgeries, encompassing both partial and complete joint replacements. It accommodates four procedures each day from Monday to Friday, with the exception of Wednesdays. The procedural workflow comprises six stages: Anesthesia Preparation, patient positioning, the surgical procedure itself, patient exiting the room and turnover, and the final stage, as illustrated in Figure 1 (Al Zoubi et al., 2022). The scientific names and abbreviations for these stages are provided in Table 1.

A successful day in this arthroplasty operating room is defined as completing all four procedures within the allocated time span of 8 h, which falls between 7:30 am and 3:30 pm. The Surgical Success Rate (SSR) serves as the metric to gauge the percentage of successful days. The initial SSR was notably low at 39%, contributing to an annual overtime cost of ∼$570,000 for our institution.

2.2 Previous work

Recently, we proposed a comprehensive solution to address this challenge, a data-driven, Machine Learning (ML)-based, prescriptive analytics system (PAS). This decision support system not only predicts the likelihood of a particular day being successful based on temporal variables but also monitors each stage of the procedure in real-time. It adjusts its predictions as needed

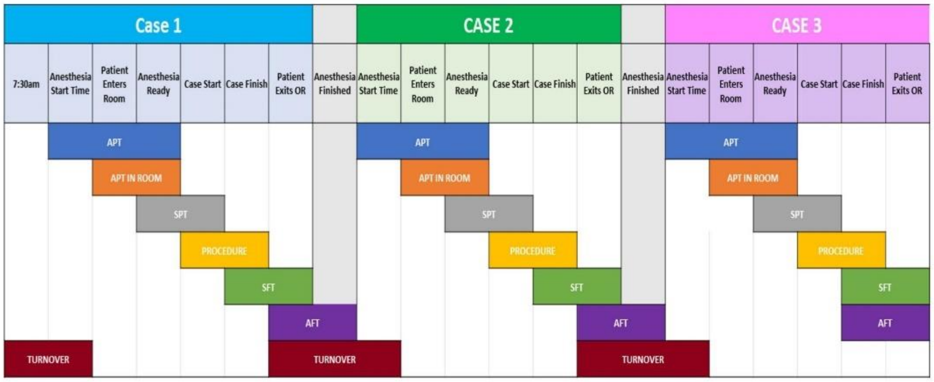


FIGURE 1
A breakdown of each step-in arthroplasty, along with the transition between two stages (turnover), is outlined. The upper section indicates the stage markers for a case. Only showing three out of four cases a day.

and provides actionable suggestions at each stage to enhance the likelihood of success (Al Zoubi et al., 2022).

The journey toward creating the PAS, aimed at enhancing the Surgical Success Rate (SSR) of high-volume arthroplasty surgeries, began by implementing Descriptive Analytics on retrospective data of ∼5,000 surgeries (Al Zoubi et al., 2024). In addition to providing actionable insights, this phase allowed us to categorize the surgical recorded parameters into patient metrics, team metrics, and time metrics. For each group of metrics, we developed a framework that serves as a decision support system to enhance SSR independently.

As the fundamental research progressed into the Predictive Analytics stage, we proposed the adoption of a Workflow/Time monitoring framework (WTMF; Al Zoubi et al., 2022), for the transformation of our PAS. The WTMF effectively aligns with our objective of improving SSR through improving the overall team efficiency. The conversion of WTMF into a prescriptive system involved several stages, such as conducting what-if scenarios, generating benchmarks, and identifying a set of actions and recommendations, which were determined through multidisciplinary positive deviance seminars (Gold et al., 2022). The Machine Learning (ML) engine in our Predictive Analytics System (PAS) predominantly relies on the decision tree technique. This is true for both predicting the Surgical Success Rate (SSR) and generating dynamic benchmarks. Additionally, it has the capability to incorporate various other supervised learning classifiers specifically for SSR prediction. The imperative for an interpretable ML model is highlighted, particularly in the context of benchmark generation. We elucidated and compared several ML techniques in our previous work to underscore this necessity.

TABLE 1 Medical names and acronyms for arthroplasty phases.

Time metrics
Anesthesia preparation time (APT)
Anesthesia finish time (AFT)
Surgical preparation time (SPT)
Surgery finish time (SFT)
Turnover
Surgery (procedure) time

Ultimately, working under the assumption of utilizing the WTMF, we conducted calculations to determine both cost savings and the possibility of adding an extra 5th joint surgery on days when the initial four joint surgeries were successfully completed on time, all in an 8-h window (Al Zoubi et al., 2023).

The central focus of this study revolves around the comparison between theoretical concepts and practical outcomes. We assess the performance of our algorithms and evaluate the efficacy of implementing the WTMF into an integrated PAS on prospective patient surgeries.

The overall journey of designing and implementing our AI-driven solution to improve the SSR is summarized in Figure 2. The boxes below display the titles of published articles linked to each respective phase, identified by their corresponding reference numbers in the bibliography.

2.3 Validation on arthroplasty surgeries—patient dataset

The PAS was implemented and validated at TOH's Riverside Campus, involving a team of seven arthroplasty surgeons, along with a group of nurses and anesthesiologists. The surgical

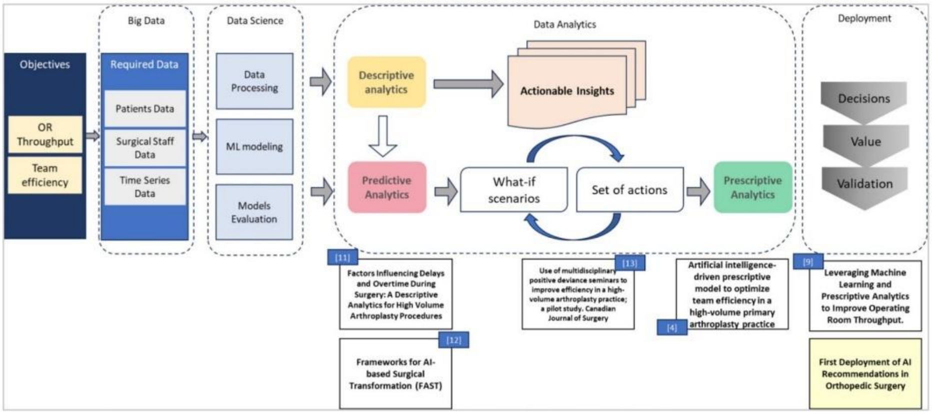


FIGURE 2
Design and deployment roadmap of AI-driven recommendations to improve the surgical success rate.

TABLE 2 Statistical summary of the 228 patient surgeries used in the validation of the PAS.

Surgeries	228
Number of surgeons	7
Number of nurses team	43
Number of anesthesiologist	13
Female patients	130
Male patients	98
Average age	65.3 ± 8.4
Average Body Mass Index (BMI)	28.7 ± 5.5
Average American Society of Anesthesiologists classification (ASA)	2 ± 0.7
Average number of surgeries/OR/day	4.95 ± 0.6

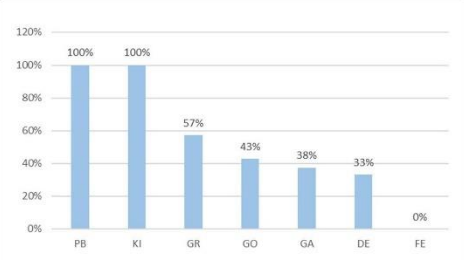


FIGURE 3
SSR per surgeon.

TABLE 3 Central tendencies for benchmarks observed on Saturday surgeries.

Metric	APT	SPT	Procedure	AFT	Turnover
Minimum	01:00	00:00	30:00	01:00	02:00
Maximum	41:00	26:00	04:00	28:00	45:00
Mean	11:45	09:05	56:26	04:05	20:47
Median	11:00	09:00	55:00	04:00	20:00
Standard deviation	0.4%	0.3%	1.5%	0.2%	0.5%

procedures were conducted weekly, utilizing two operating rooms each week, spanning ~23 Saturdays in 2023. Table 2 offers a concise overview of crucial statistical data sourced from 228 patients who underwent hip and knee replacement surgeries.

Additionally, the surgical team received pre-surgery targeted benchmarks and recommendations like the ones in our previous work 4. Further elaboration on the dataset and its potential impact on the outcomes is provided below.

2.3.1 Patient age, BMI, and gender

The demographic characteristics of the dataset reveal that the surgeons performed operations on individuals of advanced age and with high BMI. The results of descriptive analytics indicated that age and BMI have minimal influence on the SSR (Al Zoubi et al., 2024). However, in our Saturday data, a notable gender disparity was observed, with surgeons operating on ~25% more females than males. This discrepancy is expected to have an impact on achieving a higher SSR, as females tend to have a shorter operating time by ~5 min when compared to males. For this dataset, this observation remains consistent with our previously discussed conclusions in Al Zoubi et al. (2024).

2.3.2 Patient ASA

Performing surgeries on patients with pre-existing medical conditions raises the likelihood of complications and consequently, the potential delay of the 4th joint surgery. The Saturday data reveals an average ASA score of 2, with a standard deviation of 0.7, indicating that many patients have underlying medical conditions.

2.3.3 Surgeon demographics

The quantity of surgeries performed by each surgeon ranges from a maximum of 40 surgeries to a minimum of 29 per surgeon. The SSR for each surgeon is presented in Figure 3.

Furthermore, we conclude that the surgeon's experience does not exhibit a direct correlation with SSR. For instance, even though "GO" possesses over a decade of experience in hip surgeries, his SSR is only 43%. This finding aligns with the conclusion reached in our earlier publication (Al Zoubi et al., 2024).

2.4 Outcomes and observations

Table 3 showcases the central tendency values for the benchmark metrics observed during the 23 Saturdays.

To facilitate comparisons, we have included the mean values of these metrics in Table 4, derived from the earlier generated machine learning output (MLO) benchmarks in Al Zoubi et al. (2023). When comparing with the chosen benchmarks, it becomes evident that there is a notable improvement in SSR. The calculated SSR for Saturday surgeries reached 93%, falling between the MLO values of 91 and 100%. Interestingly, it appears that a procedure that is 3½ min faster has a more substantial impact on SSR than the cumulative 2-min slowdown in both turnover and APT when compared to MLO-Fast Performance. This supports our earlier conclusion in Al Zoubi et al. (2022), indicating that a 1-min difference in a specific stage can have a significantly larger impact on the outcome compared to a 1-min difference in another stage due to non-linearity.

It's crucial to keep in mind that before the implementation of the PAS, the overall SSR at The Ottawa Hospital consistently remained below 39% for arthroplasty surgeries, regardless of the surgeon involved, and including the positive deviance surgeon whose SSR was 68% (Jung et al., 2020). However, with the implementation of PAS, the SSR has surged to 93%. This achievement significantly surpasses the suggested Machine Learning Output (MLO) baseline of 77%, not to mention the pre-PAS SSR, which stood at a mere 39%.

TABLE 4 Machine learning output (MLO) benchmarks compared to Saturday surgery practices.

Benchmarks scenario	APT	SPT	Procedure	AFT	Turnover	SSR
MLO-optimal performance	7:00	7:00	62:30	7:00	20:00	100%
Saturday clinical practices	11:45	9:05	56:26	4:05	20:47	**93%**
MLO-fast performance	10:00	12:00	60:00	4:00	20:00	91%
MLO-baseline	10:30	20:00	71:30	20:30	21:30	77%
MLO-fair performance	18:30	20:00	71:30	20:30	21:30	64%
MLO-poor performance	18:00	20:00	77:00	8:00	31:00	0%

Bold value indicates the achieved SSR value leveraging the PAS system.

3 Discussion

This Community Case Study aimed to evaluate the actual improvement of TOH's arthroplasty operating room efficiency; thus, we have opted to assess and validate the performance of our designed prescriptive analytics system originating from one of the three AI-driven frameworks we had previously published, namely the WTMF.

The initial implementation of the PAS occurred at TOH and involved a collective effort from surgeons, nurses, and anesthesiologists who voluntarily chose to extend their work to Saturdays, aligning with the Ontario government's initiative to increase surgical capacity. TOH has allocated two operating rooms specifically for hip and knee replacement surgeries on Saturdays.

We had anticipated that upon the initial implementation of the WTMF, the surgical team would strive to reach the recommended baseline SSR of 77%. To our surprise, the team achieved a 93% success rate in completing their fourth joint case within an 8-h window. Furthermore, they were able to accommodate an additional case 52% of the time within the same 8-h window (Al Zoubi et al., 2023). Nevertheless, certain considerations must be considered regarding our results.

First, performing surgeries on weekends and specifically on Saturdays, offers the advantage of minimal interruptions from other hospital entities or phone calls, creating a conducive environment for surgeries. This could have had a positive impact on the outcomes. However, in our previous study in Al Zoubi et al. (2024), we demonstrated that the day of the week itself influences SSR. We found that SSR tends to be lower the closer the surgical day is to the weekend (since people look forward to ending their week quickly). This contrast in the impact of the day of the week, with potential positive and negative effects, makes it challenging to definitively determine its true influence on SSR. However, the day of the week does indeed have a distinct impact compared to surgeries conducted on weekdays.

The second factor pertained to calculating turnover times for the first case. During regular workweeks, nurses' shifts commence at 7:30 am, and the time of the first patient's entry into the operating room varies depending on team efficiency. It is well-established that the earlier the patient enters the room, the sooner the day concludes. However, on Saturdays, due to the absence of nursing union policies dictating start times, surgeons have the flexibility to commence surgeries at their discretion. To ensure data comparability, the 8-h time stamp needed to commence when the patient entered the room, as the time prior to the entry of the first case could fluctuate based on the "shift start" time.

The third consideration arises from the fact that one of the surgeons had previously been identified as an individual demonstrating positive deviance (PD) during the seminars. This prior acknowledgment could potentially introduce a positive bias to the outcomes. However, the results indicated that another surgeon achieved the same SSR as the PD surgeon, leaving us uncertain as to whether the PD recognition had influenced the data or if this level of success could be achieved under normal circumstances. It's important to note that there were no positive deviance seminars conducted specifically for this group of surgeons on Saturdays; the established benchmarks were simply communicated amongst team members as well as the culture of working together.

The fourth challenge we encountered was the manual implementation of the PAS, which involved fixed benchmark values being communicated to the practitioners. In case of any delays, surgeons had to refer to suggested actions from previous positive deviance seminars to manage the subsequent stages. Conversely, if we are to transform the PAS into a real-time decision support software, it will have the capability to generate adjusted benchmarks for upcoming stages through AI, which can assist in completing them on time or even ahead of schedule. The anticipated flexibility provided by the real-time PAS is expected to yield superior outcomes.

Finally, as explained in the preceding section there are 25% fewer male patients compared to female patients. This factor may have had a positive effect on the outcomes.

4 Conclusions

The transformed WTMF into PAS was put into action within a real clinical setting for validation purposes. This implementation spanned 228 surgical cases, demonstrating the effectiveness of the algorithms, resulting in improved team efficiency, and increased operating room throughput. This is evident through the attainment of a 93% SSR and the ability to accommodate an additional case every other week, all without incurring any additional costs.

To our knowledge, this is the first experience in deploying AI recommendations in orthopedic surgery. The work done to improve SSR in arthroplasty surgeries can be extended to other types of high-volume surgeries as well.

Two alternative frameworks previously published Surgical Team Scheduling Framework (STSF) and Patients Schedular

Framework (PSF), can likewise be converted into PAS using the same techniques and methodologies employed for the WTMF. This transformation can be accomplished by creating what-if scenarios, gathering actions for each scenario, and leveraging both multidisciplinary and interdisciplinary approaches. However, the implementation of the STSF may be controversial since at its premise, the AI recommendation is demanding specific individuals to possibly compromise in certain work activities to increase in the SSR.

Lastly, we have the desire to investigate more critical and detailed factors, such as the behavior of the staff within the operating room using cameras (Jung et al., 2020), or the potential impact of the machine and instrument vendors on operational delays. Additionally, we would like to analyze data related to drugs and anesthesia types to understand their impact on early wake-up occurrences. Another possibility is merging multiple frameworks. However, the availability of data becomes a limiting factor in exploring these aspects thoroughly.

Data availability statement

Requests to access these datasets should be directed to falzo100@uottawa.ca.

Ethics statement

The studies involving humans were approved by Ottawa Health Science Network Research Ethics Board (OHSN-REB). The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

FA: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. KK: Data curation, Writing – review & editing. PB: Conceptualization, Writing – review & editing. PF: Conceptualization, Methodology, Project administration, Resources, Supervision, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

References

Al Zoubi, F., Beaulé, P. E., and Fallavollita, P. (2024). Factors influencing delays and overtime during surgery: a descriptive analytics for high volume arthroplasty procedures. *Front. Surg.* 10:1242287. doi: 10.3389/fsurg.2023.1242287

Al Zoubi, F., Gold, R., Poitras, S., Kreviazuk, C., Brillinger, J., Fallavollita, P., et al. (2022). Artificial intelligence-driven prescriptive model to optimize team efficiency in a high-volume primary arthroplasty practice. *Int. Orthop.* 22:5475. doi: 10.1007/s00264-022-05475-1

Al Zoubi, F., Khalaf, G., Beaulé, P. E., and Fallavollita, P. (2023). Leveraging machine learning and prescriptive analytics to improve operating room throughput. *Front. Digit. Health* 5:1242214. doi: 10.3389/fdgth.2023.1242214

Bartek, M. A., Saxena, R. C., Solomon, S., Fong, C. T., Behara, L. D., Venigandla, R., et al. (2019). Improving operating room efficiency: machine learning approach to predict case-time duration. *J. Am. Coll. Surg.* 229, 346–354e3. doi: 10.1016/j.jamcollsurg.2019.05.029

Bertsimas, D., Pauphilet, J., Stevens, J., and Tandon, M. (2022). Predicting inpatient flow at a major hospital using interpretable analytics. *Manufact. Serv. Operat. Manag.* 24:971. doi: 10.1287/msom.2021.0971

Bohr, A., and Memarzadeh, K. (2020). The rise of artificial intelligence in healthcare applications. *Artif. Intell. Healthc.* 2, 25–60. doi: 10.1016/B978-0-12-818438-7.00002-2

Erekat, A., Servis, G., Madathil, S. C., and Khasawneh, M. T. (2020). Efficient operating room planning using an ensemble learning approach to predict surgery cancellations. *IISE Trans. Healthc. Syst. Eng.* 10, 18–32. doi: 10.1080/24725579.2019.1641576

Gold, R., Al Zoubi, F., Brillinger, J., Kreviazuk, C., Garvin, D., Schramm, D., et al. (2022). Use of multidisciplinary positive deviance seminars to improve efficiency in a high-volume arthroplasty practice; a pilot study. *Can. J. Surg.* 66, E1–E7. doi: 10.1503/cjs.018121

Hu, Y., Chan, C. W., and Dong, J. (2021). Prediction-driven surge planning with application in the emergency department. *Manage Sci.* 75.

Jiang, L., Wu, Z., Xu, X., Zhan, Y., Jin, X., Wang, L., et al. (2021). Opportunities and challenges of artificial intelligence in the medical field: current application, emerging problems, and problem-solving strategies. *J. Int. Med. Res.* 49:157. doi: 10.1177/03000605211000157

Jung, J. J., Jüni, P., Lebovic, G., and Grantcharov, T. (2020). First-year analysis of the operating room black box study. *Ann. Surg.* 271, 122–127. doi: 10.1097/SLA.0000000000002863

Maier-Hein, L., Eisenmann, M., Sarikaya, D., März, K., Collins, T., Malpani, A., et al. (2022). Surgical data science - from concepts toward clinical translation. *Med. Image Anal.* 76:102306. doi: 10.1016/j.media.2021.102306

Nwoye, E., Woo, W. L., Gao, B., and Anyanwu, T. (2023). Artificial intelligence for emerging technology in surgery: systematic review and validation. *IEEE Rev. Biomed. Eng.* 16:3183852. doi: 10.1109/RBME.2022.3183852

Safavi, K. C., Khaniyev, T., Copenhaver, M., Seelen, M., Langle, A. C. Z., Zanger, J., et al. (2019). Development and validation of a machine learning model to aid discharge processes for inpatient surgical care. *J. Am. Med. Assoc. Netw. Open* 2:17221. doi: 10.1001/jamanetworkopen.2019.17221

Schiele, J., Koperna, T., and Brunner, J. O. (2021). Predicting intensive care unit bed occupancy for integrated operating room scheduling via neural networks. *Naval Res. Logist.* 68, 65–88. doi: 10.1002/nav.21929

Strömblad, C. T., Baxter-King, R. G., Meisami, A., Yee, S. J., Levine, M. R., Ostrovsky, A., et al. (2021). Effect of a predictive model on planned surgical duration accuracy, patient wait time, and use of presurgical resources. *J. Am. Med. Assoc. Surg.* 156:315. doi: 10.1001/jamasurg.2020.6361

Chapter 8: Thesis Conclusion

The transformed WTMF into PAS was put into action within a real clinical setting for validation purposes. This implementation spanned across 228 surgical cases, demonstrating the effectiveness of the algorithms, resulting in improved team efficiency, increased operating room throughput, and a cost-effective solution. This is evident through the attainment of a 93% SSR and the ability to accommodate an additional case every other week, all without incurring any additional costs.

The work done to improve SSR in arthroplasty surgeries can be extended to other types of high-volume surgeries as well. There is ongoing work to collect retrospective data for General Surgery at The Ottawa Hospital. The aim is to apply the fundamental research published in our journals to the General Surgery scenario.

The two alternative frameworks, STSF and PSF, can likewise be converted into PAS using the same techniques and methodologies employed for the WTMF. This transformation can be accomplished by creating what-if scenarios, gathering actions for each scenario, and leveraging both multidisciplinary and interdisciplinary approaches just as we carried out in Chapter 3. However, the implementation of the STSF may be controversial since at its premise, the AI recommendation is demanding specific individuals to work together to guarantee an increase in the SSR.

Lastly, I have a desire to investigate more critical and detailed factors, such as the behavior of the staff within the operating room using cameras, or the potential impact of the machine and instrument vendors on operational delays. Additionally, I would like to analyze data related to drugs and anesthesia types to understand their impact on early wake-up occurrences. Another possibility is merging multiple frameworks together. However, the availability of data becomes a limiting factor in exploring these aspects thoroughly.

Appendixes

Appendix A

Statistical Analysis

The table below illustrates the correlation between each metric and the SSR, which were the subjects of comparison in the initial stage of our analyses. The correlation values closely align with the findings depicted in graphs and visually described in our published journal article. None of the metrics exhibit a strong, direct correlation with the SSR.

Table A: Metrics vs SSR correlations

Metric	Pearson Correlation With SSR	Spearman Correlation With SSR
Age	-0.088	-0.092
ASA	-0.038	-0.046
Turnover	-0.079	-0.069
APT	-0.181	-0.195
AFT	-0.067	-0.085
APTinRoom	-0.094	-0.109
BMI	0.018	0.004
SFT	-0.107	-0.124
SPT	-0.106	-0.109
Procedure	-0.150	-0.197
Sex	-0.031	-0.031

It is crucial to note that certain phrasings were modified in the published version. Our objective was not to draw conclusions based on the data, as the task primarily involved a descriptive rather than a statistical analysis.

For example, our focus was on understanding why female surgeries tended to require less time than male surgeries, rather than delving into the statistical significance of this observation. We relied on referenced studies in the article and expert opinions to support the belief that gender differences in operation time play a role in SSR, a judgment that was subsequently validated in chapters 4, 5, and 7. Another example, the "Campus" metric was not subjected to analysis due to its low sample size of 1:7 in the civic category, which constituted less than 1% of our 5k data, leading to its exclusion from further consideration. The Fisher exact test statistics for the two aforementioned examples were **0.3411** and **0.1839**, indicating a lack of statistical significance to draw conclusions regarding the importance to the SSR at this stage of our study.

Two additional considerations are addressed in this appendix. The first pertains to how we computed the SSR per gender, which was done as flat ratios. For instance, we examined all male surgeries, dividing success days by the total number of days, and similarly for female surgeries. Patients were randomly assigned to operating rooms, and there were no scheduling techniques, such as ensuring a mix of male and female patients on specific days.

The second concern addresses why we did not incorporate additional data about the surgeons in our analysis, such as their years of experience, educational backgrounds, and other relevant factors. Unfortunately, we lacked this information and did not believe it was available in our institution in a retrievable format.

Appendix B

Prescriptive Analytics System (PAS) building Blocks

Our proposed resolution to the identified issue of low Operating Room throughput initiates with data and culminates in a Prescriptive Analytics System (PAS), which serves as a decision support system. We employed data to construct multiple machine-learning models, utilizing these models as engines within our frameworks. Subsequently, we refined our frameworks to function as a Prescriptive Analytics System.

Our goal is to assist diverse healthcare institutions and professionals in identifying the most suitable optimization framework for their needs. These institutions rely on machine learning (ML) models with varying data input streams, creating a versatile set that caters to a wide range of healthcare settings. This flexibility allows institutions to begin with the framework that aligns with the data they currently possess. Even if a healthcare institution has access to only one aspect of data, such as patient metrics or specific steps in surgical procedures, they can adopt one or multiple of our frameworks. The AI-driven frameworks we propose have been meticulously designed and validated using patient-specific orthopedic data.

This section provides a comprehensive elucidation of each component of our proposed product, **the PAS**, including the input, output, building modules, and the necessary elements for each component to constitute a fully functional PAS.

Diagram B-1 illustrates the fundamental modules involved in constructing the PAS. The process initiates with the **Data Module**, followed by the **Framework Module** and ending by the **Decision Support System (DSS) module**.

The remainder of Appendix B delves into specific instances of input and output for each module/component within the PAS.

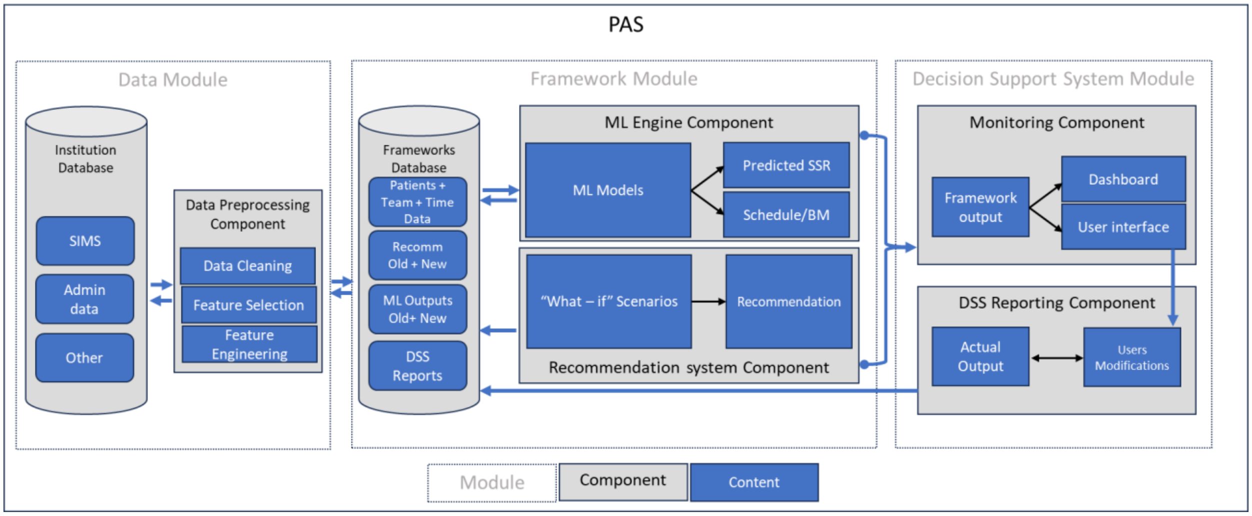


Diagram B-1: PAS Components and building blocks

1.1 Data Module (DM)

The Data Module comprises two primary components. Firstly, there is the institution's database containing all the necessary data for any of our frameworks, such as Surgical Information Management Systems, which provides information about patients and time metrics. The second component is the data preprocessing component. Its role is to extract only the essential data for each framework, clean it, and perform feature engineering when necessary. The output of the data module is refined, ready-to-use, and selected metrics for each of the proposed frameworks: STSF, PSF, and the WTMF.

1.2 Framework Module (FM)

Once the Data Module (DM) completes the data preparation, it is subsequently sent to the Frameworks Module, which consists of three components: the ML Engine, the framework database, and the recommendation system.

(a) The Machine learning engine component

The Machine learning engine component houses all the machine learning models necessary for each framework. The output of these models includes the predicted SSR and the essential metric values required to attain a specific SSR.

In Chapter 3, we provided a detailed explanation of how we conducted comparisons for our models within each framework. While selecting several evaluation metrics, we accorded greater importance to cross-validation accuracy and the overfitting metric. To tune hyperparameters, we employed the grid search option, resulting in the outcomes showcased in Table 8 in Chapter 3 Three which was updated in this appendix below. The parameters of the estimator underwent optimization through cross-validated grid search across a parameter grid. The chosen hyperparameters for each model within the framework are detailed in the table found in the Appendix of Chapter 3 “Appendix of the published article”. Additionally, we illustrated how altering a specific metric value can influence the SSR, in other words, the OR throughput for each framework.

In this appendix, we present an illustration of the anticipated outcomes of machine learning model and offer guidance on leveraging these outputs within the framework to achieve the goal of performing four arthroplasties a day for the WTMF. It is important to note that the other two frameworks, namely the PSF and the STSF, are beyond the scope of this work. As we have not developed the other components, other than the ML, of the framework module, such as the DSSM and the recommendation systems. Providing a detailed example for these frameworks would be challenging. Nevertheless, a schedule for one week of patients and how it is organized is included, albeit without specific recommendations.

Workflow/Time Monitoring Scenario

In this section, I will elucidate the functioning of ML input and output for a complete day comprising four arthroplasty surgeries within our WTMF scenario. The necessary timestamps for calculating time metrics are provided in Table B-1 below. The responsibility of generating the required time metrics as input for the ML models lies with the Data Preprocessing component in the Data module. The calculation occurs each time there is an updated value for any of the timestamps. The system's default values are set as our baseline values, corresponding to the 77% SSR benchmarks explained in Chapter 5.

The day is ideally scheduled to initiate its first surgery, "case 1," at 7:30 am with a turnover time of 0 minutes. If the day commences at a different time, for instance, 7:45 am, the preprocessing module calculates the turnover value as 15 minutes (7:45 -7:30). The formula below is employed to estimate and calculate the remaining time metrics of the case. $\boldsymbol{Metric_{new}}$ represents the estimated time for the new metric. The term "Delay or gain" signifies the difference between the actual metrics value once the stage is completed and the baseline value of that metric. If the actual value is less than the baseline value, it is considered a gain, and its sign in the formula becomes a negative value. Conversely, if the actual metric value exceeds the baseline value, it is regarded as a delay.

$$\boldsymbol{Metric_{new} = \ Metric\ _{baseline} + \frac{(Delays\ or\ gains)_{last\ stage}}{Number\ of\ Metrics_{New}}}$$

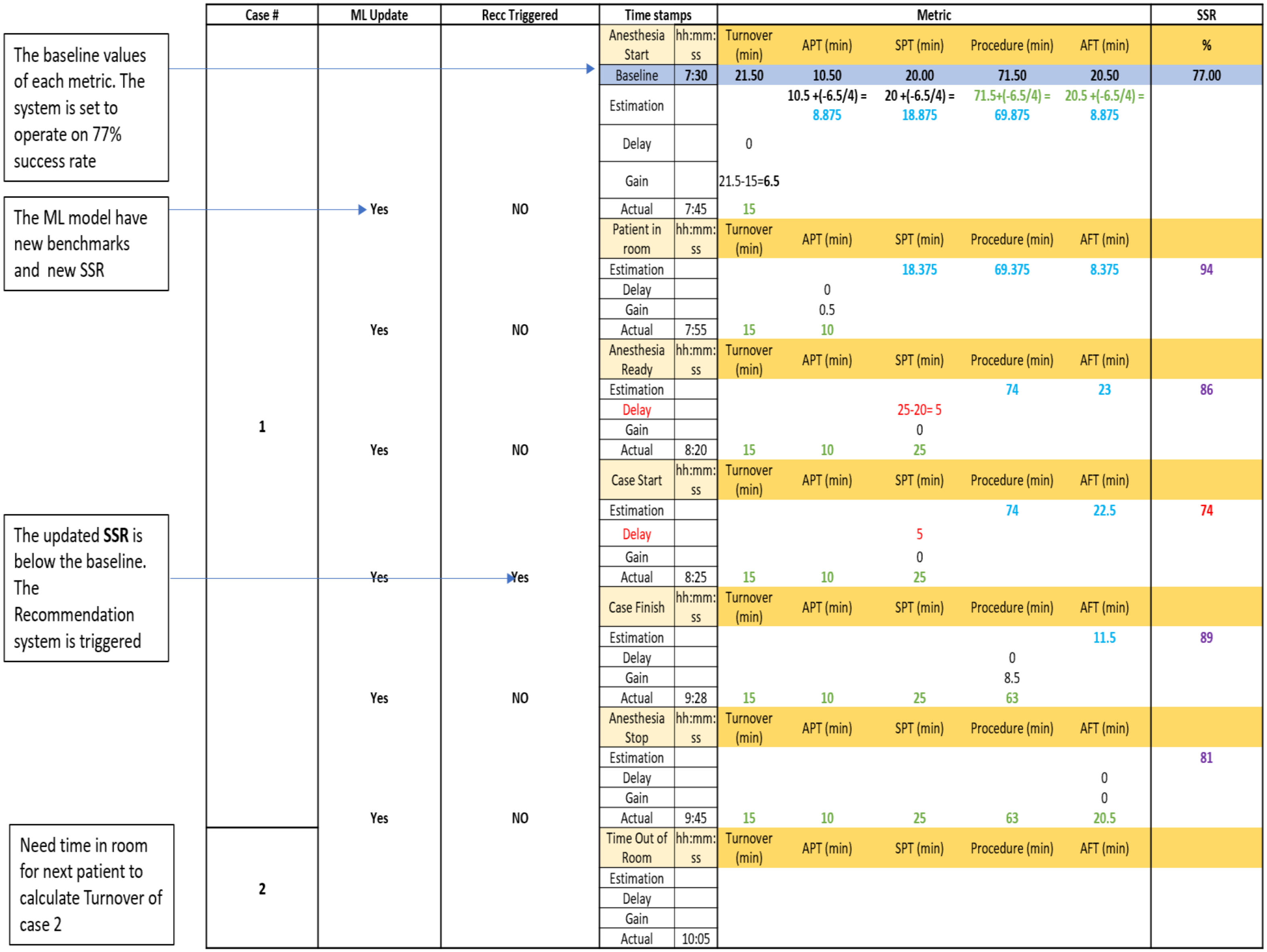
For the given example, with an actual turnover value of 15 minutes and a baseline value of 21.5 minutes, we experience a gain in time of 6.5 minutes ("-6.5"). The subsequent metrics will then have their estimated time as their baseline values minus 6.5/4, which is 1.625 minutes, as outlined in diagram B-2. The assumption here is that delays and/or gains in time from the previous stage are evenly distributed among the upcoming stage.

The actual metric values are calculated at the conclusion of each stage once the required timestamp to calculate the stage is available. The ML model's input is then updated with the actual metrics values for completed stages and with the estimated values for metrics where actual values are not yet available. A SSR is generated at each new model run.

In the B-2 diagram, we presented the initial case for a surgical day, highlighting instances of when and how the system updates the ML input, ML output, and the predicted SSR. The calculation formula is displayed in the first row to illustrate the application of the new metric formula. Additionally, it outlines specific points where the system activates the recommendation system to address potential delays. Metrics displayed in blue and green (actual values and estimated values) font serve as the ML model input and are updated with each new timestamp. The SSR is also updated whenever there is new ML input. The recommendation system is triggered only when the estimated SSR falls below the baseline value. This sequence repeats in the diagram until the completion of all four surgeries.

Table B-1: Time stamps to calculate the Time Metrics

Time Stamps	**Time Metrics (Intervals)**
Date	Turnover
Case no	Anesthesia Preparation time (APT)
Time in Room	Surgical Time (Case or Procedure)
Anesthesia Start	Anesthesia Finish Time (AFT)
Anesthesia Ready	Surgical Preparation Time (SPT)
Anesthesia Stop	APT in room
Case Start	Surgery finish time (SFT)
Case Finish	
Time Out of Room	



The baseline values of each metric. The system is set to operate on 77% success rate

The ML model have new benchmarks and new SSR

The updated **SSR** is below the baseline. The Recommendation system is triggered

Need time in room for next patient to calculate Turnover of case 2

Case #	ML Update	Recc Triggered	Time stamps		Metric					SSR
1			Anesthesia Start	hh:mm:ss	Turnover (min)	APT (min)	SPT (min)	Procedure (min)	AFT (min)	%
			Baseline	7:30	21.50	10.50	20.00	71.50	20.50	77.00
			Estimation			10.5 +(-6.5/4) = 8.875	20 +(-6.5/4) = 18.875	71.5+(-6.5/4) = 69.875	20.5 +(-6.5/4) = 8.875	
			Delay		0					
			Gain		21.5-15=6.5					
	Yes	NO	Actual	7:45	15					
			Patient in room	hh:mm:ss	Turnover (min)	APT (min)	SPT (min)	Procedure (min)	AFT (min)	
			Estimation				18.375	69.375	8.375	94
			Delay			0				
			Gain			0.5				
	Yes	NO	Actual	7:55	15	10				
			Anesthesia Ready	hh:mm:ss	Turnover (min)	APT (min)	SPT (min)	Procedure (min)	AFT (min)	
			Estimation					74	23	86
			Delay				25-20= 5			
			Gain				0			
	Yes	NO	Actual	8:20	15	10	25			
			Case Start	hh:mm:ss	Turnover (min)	APT (min)	SPT (min)	Procedure (min)	AFT (min)	
			Estimation					74	22.5	74
			Delay				5			
			Gain				0			
	Yes	Yes	Actual	8:25	15	10	25			
			Case Finish	hh:mm:ss	Turnover (min)	APT (min)	SPT (min)	Procedure (min)	AFT (min)	
			Estimation						11.5	89
			Delay					0		
			Gain					8.5		
	Yes	NO	Actual	9:28	15	10	25	63		
			Anesthesia Stop	hh:mm:ss	Turnover (min)	APT (min)	SPT (min)	Procedure (min)	AFT (min)	
			Estimation							81
			Delay						0	
			Gain						0	
	Yes	NO	Actual	9:45	15	10	25	63	20.5	
2			Time Out of Room	hh:mm:ss	Turnover (min)	APT (min)	SPT (min)	Procedure (min)	AFT (min)	
			Estimation							
			Delay							
			Gain							
			Actual	10:05						

Diagram B-2: Detailed example of how the ML models updates the benchmarks and the SSR

Patient Scheduling Scenarios

Table B-2 below displays the scheduling of four patients per day along with the predicted Surgical Start Rate (SSR) for each day. The scheduler incorporates patient metric conditions to facilitate flexibility and ease of selection based on both patient and surgeon availability. For example, on Day One, the schedule prioritizes the booking of four healthy women under the age of 42 for hip arthroplasty procedures. In this scenario, the likelihood of completing the fourth surgery before 3:30 pm is notably high, with a probability of 91%. On Day Two, even when considering a similar group of patients in poorer health conditions (ASA 3 and 4), the SSR remains remarkably high at 86%.

In the context of the PSF, we provided an example of patient scheduling using the decision tree model, acknowledging that it was not the optimal model, as discussed in Chapter 3. Nevertheless, our aim was to demonstrate how patients could potentially be scheduled based on metric values. An example of a recommendation system in this scenario could involve suggesting changes, such as replacing a male with a female patient on a specific day to enhance the likelihood of SSR or choosing a patient with a different type of surgery.

However, it's important to note that ML models are generally not easily explainable, and we may only have a schedule of patients and the predicted SSR without precise insights into the reasoning behind the recommendations. This complexity can make it challenging to select an appropriate recommendation system, potentially leading to scenarios where a recommendation system unrelated to patient metrics is chosen. For instance, it's conceivable that our WTMF might function as a recommendation system for the PSF, showcasing the intricacies of working with machine learning models in practice.

Table B-2: One week schedule of patients

Day	Surgery	Anesthesia	Gender	Age	ASA	SSR
1	HR, THA	Spinal	4F	<42	1,2	91%
2	HR, THA	Spinal	4F	<42	3,4	86%
3	UKA	Spinal	3F, 1M	<57	Any	67%
4	UKA	Spinal	2F,2M	58-70	Any	64%
5	THA	Spinal	4M	42-53	1,2	58%
6	THA	General	4M	42-53	1,2	54%

(b) Recommendation System Component

The recommendation system specifically designed for the WTMF, was created using a distinctive approach known as the positive deviance (PD) seminar. Introducing any other method for implementing a recommendation system could contribute to transforming the framework into a prescriptive analytics decision support system.

The decision to employ PD seminars as the source of recommendations for our framework, WTMF, was in line with our objective of enhancing operating room throughput by improving team efficiency. While alternative approaches, such as acquiring additional resources, hiring consultancy services, or adding personnel to oversee surgery flow, were available, they would have incurred extra costs. Moreover, our intention was to encourage the team's self-improvement by setting a positive example within the existing team dynamics and facilitating self-monitoring based on agreed-upon achievable and feasible performance standards.

The positive deviance (PD) seminar, functioning as a recommendation system, underwent testing and validation independently of the machine learning engine. To illustrate, we implemented these recommendations in a 57-day trial focused on arthroplasty, and the outcomes are detailed in the Table B-3 below.

Table B-3: Walsh test for the PD seminar recommendation implementation

Metric	APT	APT_IN_ROOM	SPT	CASE	SFT	AFT	Turnover	**Success**
Mean-Before	40:54	16:28	12:07	1:03:36	04:50	05:44	15:00	**14%**
Mean-After	36:48	12:28	12:40	59:51	05:12	07:21	18:51	**64%**
STD-Before	13:20	09:11	06:04	08:22	01:59	03:05	33:13	0.3499
STD-After	11:08	10:01	07:21	12:18	02:20	03:52	49:06	0.229592
t-stat	1.3858	1.8347	-0.2209	1.6845	-1.4749	-2.3791	-0.3872	-690.9356
p-value	0.0862	**0.0358**	0.4129	**0.0482**	0.0723	**0.0102**	0.3499	**0.0000**
DF	46	58	63	72	71	61	74	37

(a) The Framework database Component

The role of the framework database is to manage and store the processed and categorized data provided by the data module. Additionally, it serves as a repository for ML output data, what-if scenarios, selected recommendations, and the actual data received at the completion of each stage by the Decision Support System Module (DSSM).

1.3 Decision Support System Module

To transform the PAS into a fully developed product, beyond having a well-functioning framework with a tested and implemented recommendation system, additional components are necessary. These include a user interface, a dashboard, and database integration for both the input/output data of the machine learning model and the output of the system after each run.

The user interface serves to streamline interaction with the PAS system. As users utilize the dashboard to monitor the progression of surgery over time, they gain the capability to intervene in real-time. This involves allowing administrators to manually modify inputs, adjust scheduling, modify benchmarks, and either accept or reject suggestions from the system.

The comprehensive solution and the nature of the output at each stage/phase of implementation are depicted in the following diagram, B-3.

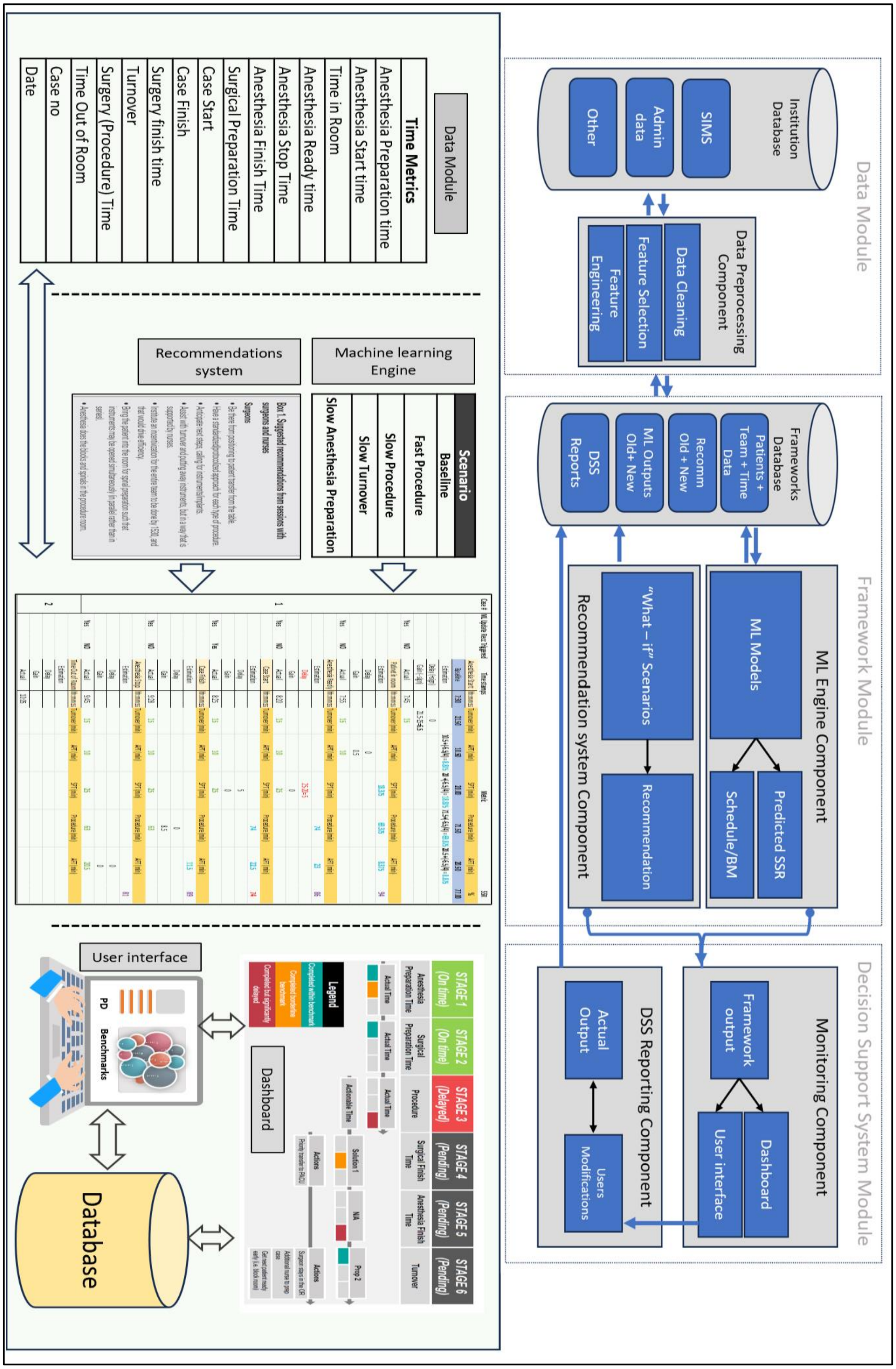


Diagram B-3: A comprehensive view of the entire PAS input/output prototype.

Table 8 Chapter 3 more details and update.

Table 8 showcases the results of each distinct ML model and facilitates model-to-model comparisons. The table employs four legends, each corresponding to a distinct color. Entries in green font denote the highest values among all ML models, specifically in terms of cross-validation accuracy metrics, while the lowest values in CV accuracy are indicated in red. Metrics highlighted in blue font are crucial considerations, particularly when weighed against other metrics, aiding in the selection of the best model for each framework. The final legend features fully highlighted rows in green, representing the chosen model for each respective framework.

As an illustration, in the context of surgical team scheduling framework, the DT model exhibited the highest cross-validation (CV) accuracy (indicated in green font), while the SVM model had the lowest (depicted in red font). The chosen model, highlighted in an entire row of green, is the XGBoost model. It was selected primarily due to its characteristics, such as being a low-overfitted model. The highest-performing model (LR) has an overfitting percentage of 11% (noted in blue).

Patient Scheduling (Numerical and Categorical)

Model	**CV Accuracy**	**AUC-ROC**	**Sensitivity**	**Specificity**	**Precision**	**Overfitting**
LR	**67%**	63%	19%	86%	57%	3%
DT	66%	65%	50%	97%	64%	1%
XGBoost	65%	92%	37%	71%	53%	5%
RF	62%	58%	36%	72%	64%	3%
DNN-ANN	62%	58%	17%	86%	56%	1%
SVM	57%	53%	9%	88%	60%	4%
Standard Deviation	0.04	0.14	0.14	0.10	0.04	0.02

Surgical team Scheduling (Categorical)

Model	**CV Accuracy**	**AUC-ROC**	**Sensitivity**	**Specificity**	**Precision**	**Overfitting**
DT	**85%**	93%	86%	80%	82%	**11%**
XGBoost	80%	92%	83%	79%	79%	3%
RF	74%	92%	85%	79%	81%	6%
SVM	71%	88%	82%	81%	80%	1%
DNN-ANN	62%	50%	49%	68%	59%	20%
LR	59%	54%	0%	100%	31%	2%

Standard Deviation	0.10	0.20	0.34	0.10	0.20	0.07
Workflow /Time Monitoring (Numerical)						
Model	**CV Accuracy**	**AUC-ROC**	**Sensitivity (Recall)**	**Specificity**	**Precision**	**Overfitting**
LR	**76%**	74%	74%	84%	62%	**8%**
SVM	75%	**81%**	74%	77%	**27%**	6%
RF	72%	72%	66%	59%	80%	1%
DNN-ANN	70%	74%	73%	59%	64%	4%
XGBoost	70%	76%	72%	61%	60%	4%
DT	68%	67%	74%	55%	53%	5%
Standard Deviation	0.03	0.04	0.03	0.12	0.17	0.02

Table 8: A modified version of Table 8 chapter 3. (Legend and colours modifications)

Appendix C

Selection Bias Analysis

Potential sources of selection bias arise in the selection of surgeons who will implement the PAS in clinical practice on Saturdays. Notably, half of the participating surgeons had already attended PD seminars, enhancing their awareness of the problem and its solutions, and coincidentally, these same surgeons participated in Saturday surgeries. Furthermore, one of these surgeons is identified as the positive deviance individual. An analysis was conducted by comparing data with and without the participation of those who attended PD seminars, as presented in the subsequent table. Cells highlighted in red font in the Delta row signify instances where the new team outperformed the PD team. Overall, the PD team demonstrated approximately a 2.5% improvement in SSR. This difference may have optimistically influenced the outcomes of the Saturday implementations.

Metric	APT, min	SPT, min	Procedure, min	AFT, min	Turnover, min	SSR
PD team	11:40	09:45	55:35	04:06	20:47	59.50
New team	11:40	09:38	56:09	04:04	20:53	57.00
Delta, min	00:01	00:07	00:34	00:02	00:07	2.50

Appendix D

Accommodating additional cases

As explained in the Introduction, to meet the increasing needs of the population and cut down the wait time for arthroplasties, dedicated funds for quality-based procedures were federally regulated to increase the number of surgeries performed, thereby creating four joint OR days in hospitals. In addition to the high SSR, PAS has also offered the possibility to address the waiting time and overtime pay issues. In Chapter 6 we presented two hypothetical scenarios of leveraging the PAS system to be able to fit an additional 5th case. In this context, here I will define the surgical success rate of incorporating a fifth case within an 8-hour time frame using PAS as SSR_5.

The SSR_5 was calculated to be at 52% during our Saturday surgeries. This value is notably higher compared to the 26% SSR_5 achieved by simply relying on the PAS at the baseline benchmark level. It's also a substantial jump compared to baseline with the 100% SSR_5 attainable through optimal performance. This implies that as the SSR for completing the initial four cases on time increases, it becomes more achievable to attain a higher SSR_5 and conversely, as the SSR for the first four cases decreases, it becomes more challenging to achieve a higher SSR_5. Table 5 presents a comparison between SSR and SSR_5 for both hypothetical PAS scenarios and the real-world outcomes of Saturday surgeries.

Table 5: SSR vs $\mathbf{SSR_5}$ between hypothesis and reality

Scenario	SSR	$\mathrm{SSR_5}$	**Translation of $\mathbf{SSR_5}$ into number of extra cases**
Hypothetical Scenario 1 **(Baseline performance)**	77%	26%	Once per month
Hypothetical Scenario 2 **(Optimal performance)**	100%	100%	Once per week
Saturday clinical practices **(Practical outcomes)**	93%	52%	Once every two weeks

Bibliography

1. Ferrata P, Carta S, Fortina M, et al (2011) Painful hip arthroplasty: Definition. Clinical Cases in Mineral and Bone Metabolism 8
2. Barua B, Rovere MC, Skinner BJ (2012) Waiting Your Turn: Wait Times for Health Care in Canada 2010 Report. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.1783079
3. Canadian Institute for Health Information (2022) Inpatient Hospitalization, Surgery and Newborn Statistics
4. Canadian Institute for Health Information (2020) Hip and Knee Replacements in Canada: CJRR Annual Statistics Summary, 2018–2019. Canadian Joint Replacement Registry (CJRR) Annual Report Ottawa, ON:CIHI
5. Canadian Institute for Health Information (2019) Canadian joint replacement registry annual report. Canadian Joint Replacement Registry (CJRR) Annual Report Ottawa, ON:CIHI
6. Clement ND, Scott CEH, Murray JRD, et al (2021) The number of patients "worse than death" while waiting for a hip or knee arthroplasty has nearly doubled during the COVID-19 pandemic A UK NATIONWIDE SURVEY. Bone and Joint Journal 103-B: https://doi.org/10.1302/0301-620X.103B.BJJ-2021-0104.R1
7. Winter J (2016) Patient Reported Outcome Measures (PROMs) in England – Data Quality Note Patient Reported Outcome Measures (PROMs) in England – Finalised data for April 2014 to March 2015
8. Health Canada G of C (2019) Canada's Health Care System - Canada.ca. Goverment Of Canada
9. (1983) The Canada Health Act. RNAO News 39:. https://doi.org/10.1515/9780228016335-011
10. CIHI (2020) National Health Expenditure Trends 2020. National Health Expenditure
11. Piccininni CR, Kwong M (2020) Refugee health care funding in Canada. Univ West Ont Med J. https://doi.org/10.5206/uwomj.v88i1.6182
12. Canadian Institute for Health Information (2023) Surgeries impacted by COVID-19: An update on volumes and wait times
13. Canadian Institute for Health Information (2023) Wait times for priority procedures in Canada, 2022

14. Canadian Institute for Health Information (2023) Wait Times for Priority Procedures in Canada - Data Tables
15. Government of Canada (2009) ARCHIVED - National Wait Times Initiative (NWTI)
16. Shohreh Majd; Zohreh Majd (2023) CLINICAL AND NON-CLINICAL MANAGEMENT IN HEALTHCARE: A COMPREHENSIVE OVERVIEW. Int J Adv Res (Indore)
17. C. Yeoh JMKST and LT (2018) Real-time locating systems and the effects on efficiency of anesthesiologists . J Clin Anesth Pain Manag 2, no. 1:
18. Porta CR, Foster A, Causey MW, et al (2013) Operating room efficiency improvement after implementation of a postoperative team assessment. Journal of Surgical Research 180:15–20. https://doi.org/10.1016/j.jss.2012.12.004
19. Eshghali M, Kannan D, Salmanzadeh-Meydani N, Esmaieeli Sikaroudi AM (2023) Machine learning based integrated scheduling and rescheduling for elective and emergency patients in the operating theatre. Ann Oper Res. https://doi.org/10.1007/s10479-023-05168-x
20. Samorani M, Blount LG (2020) Machine learning and medical appointment scheduling: Creating and perpetuating inequalities in access to health care. Am J Public Health 110
21. Salah H, Srinivas S (2022) Predict, then schedule: Prescriptive analytics approach for machine learning-enabled sequential clinical scheduling. Comput Ind Eng 169:. https://doi.org/10.1016/j.cie.2022.108270
22. Garg S, Sinha S, Kar AK, Mani M (2022) A review of machine learning applications in human resource management. International Journal of Productivity and Performance Management 71
23. Myszczynska MA, Ojamies PN, Lacoste AMB, et al (2020) Applications of machine learning to diagnosis and treatment of neurodegenerative diseases. Nat Rev Neurol 16
24. Maina EM, Oboko RO, Waiganjo PW (2017) Using machine learning techniques to support group formation in an online collaborative learning environment. International Journal of Intelligent Systems and Applications 9:. https://doi.org/10.5815/ijisa.2017.03.04
25. Lazebnik T (2023) Data-driven hospitals staff and resources allocation using agent-based simulation and deep reinforcement learning. Eng Appl Artif Intell 126:. https://doi.org/10.1016/j.engappai.2023.106783

26. Rudra Kumar M, Pathak R, Gunjan VK (2022) Machine Learning-Based Project Resource Allocation Fitment Analysis System (ML-PRAFS). In: Lecture Notes in Electrical Engineering
27. Dehnoei S, Sauré A, Ozturk O, et al (2022) A stochastic optimization approach for staff scheduling decisions at inpatient units. International Transactions in Operational Research. https://doi.org/10.1111/itor.13226
28. Maheshwari K, You J, Cummings KC, et al (2017) Attempted Development of a Tool to Predict Anesthesia Preparation Time From Patient-Related and Procedure-Related Characteristics. Anesth Analg 125:580–592. https://doi.org/10.1213/ANE.0000000000002018
29. Schiele J, Koperna T, Brunner JO (2021) Predicting intensive care unit bed occupancy for integrated operating room scheduling via neural networks. Naval Research Logistics (NRL) 68:65–88. https://doi.org/10.1002/nav.21929
30. Bartek MA, Saxena RC, Solomon S, et al (2019) Improving Operating Room Efficiency: Machine Learning Approach to Predict Case-Time Duration. J Am Coll Surg 229:346-354e3. https://doi.org/10.1016/j.jamcollsurg.2019.05.029
31. Erekat A, Servis G, Madathil SC, Khasawneh MT (2020) Efficient operating room planning using an ensemble learning approach to predict surgery cancellations. IISE Trans Healthc Syst Eng 10:18–32. https://doi.org/10.1080/24725579.2019.1641576
32. Frazzetto D, Nielsen TD, Pedersen TB, Šikšnys L (2019) Prescriptive analytics: a survey of emerging trends and technologies. VLDB Journal 28:. https://doi.org/10.1007/s00778-019-00539-y
33. Bhatt (Mishra) D, Naqvi S, Gunasekaran A, Dutta V (2023) Prescriptive analytics applications in sustainable operations research: conceptual framework and future research challenges. Ann Oper Res. https://doi.org/10.1007/s10479-023-05251-3
34. Lepenioti K, Bousdekis A, Apostolou D, Mentzas G (2020) International Journal of Information Management Prescriptive analytics : Literature review and research challenges. Int J Inf Manage 50:
35. Lepenioti K, Bousdekis A, Apostolou D, Mentzas G (2019) Prescriptive analytics: A survey of approaches and methods. In: Lecture Notes in Business Information Processing
36. Lepenioti K, Bousdekis A, Apostolou D, Mentzas G (2020) Prescriptive analytics: Literature review and research challenges. Int J Inf Manage 50
37. Mosavi NS, Santos MF (2020) How prescriptive analytics influences decision making in precision medicine. In: Procedia Computer Science

38. Dash S, Shakyawar SK, Sharma M, Kaushik S (2019) Big data in healthcare: management, analysis and future prospects. J Big Data 6:. https://doi.org/10.1186/s40537-019-0217-0
39. Javaid M, Haleem A, Pratap Singh R, et al (2022) Significance of machine learning in healthcare: Features, pillars and applications. International Journal of Intelligent Networks 3:. https://doi.org/10.1016/j.ijin.2022.05.002
40. Poornima S, Pushpalatha M (2020) A survey on various applications of prescriptive analytics. International Journal of Intelligent Networks 1:. https://doi.org/10.1016/j.ijin.2020.07.001
41. Luo W, Phung D, Tran T, et al (2016) Guidelines for developing and reporting machine learning predictive models in biomedical research: A multidisciplinary view. J Med Internet Res 18:. https://doi.org/10.2196/jmir.5870
42. Richthofen G von, Ogolla S, Send H (2022) Adopting AI in the Context of Knowledge Work: Empirical Insights from German Organizations. Information (Switzerland) 13:. https://doi.org/10.3390/info13040199
43. Loftus TJ, Balch JA, Ruppert MM, et al (2022) Aligning Patient Acuity With Resource Intensity After Major Surgery: A Scoping Review. Ann Surg 275
44. Goodair B, Reeves A (2022) Outsourcing health-care services to the private sector and treatable mortality rates in England, 2013–20: an observational study of NHS privatisation. Lancet Public Health 7:. https://doi.org/10.1016/S2468-2667(22)00133-5
45. Dwivedi S, Agrawal R, Misra R, et al (2022) Biotechnology in primary healthcare and hospital management. In: Biotechnology in Healthcare, Volume 2: Applications and Initiatives
46. Navita Mahajan; Namrata Pancholi; Seema Garg; Vibha Singh Smart healthcare: A study of barriers of digitalization of app based healthcare facilities towards consumer behavior. AIP Conference Proceedings 2782, 020085 (2023)
47. Chiu W-K, Fong BYF (2023) New Paradigms in the Business of Healthcare. In: Leung TCH, Chiu W-K, You CS-X, Fong BYF (eds) Environmental, Social and Governance and Sustainable Development in Healthcare. Springer Nature Singapore, Singapore, pp 65–78
48. Jiang F, Jiang Y, Zhi H, et al (2017) Artificial intelligence in healthcare: Past, present and future. Stroke Vasc Neurol 2