RISK MITIGATION FOCUSED ON SURGICAL CARE USING PROCESS IMPROVEMENT METHODOLOGIES IN RURAL HEALTH SYSTEMS

by

Nejc Sitar

A thesis submitted in partial fulfillment
of the requirements for the degree

of

Master of Science

in

Industrial and Management Systems Engineering

MONTANA STATE UNIVERSITY
Bozeman, Montana

December 2023

DEDICATION

I would like to dedicate this dissertation to my family Majda Sitar, Peter Sitar, and Rok Sitar. Words cannot describe how grateful I am for your sacrifice, support, and unconditional love that you have dedicated to me throughout my life. Thank you from the bottom of my heart!

ACKNOWLEDGEMENTS

First, I would like to thank my family for providing me with the support, resources, and opportunity to follow my dream of pursuing my education in the United States of America. I would like to express my gratitude to my committee chair, Dr. Bernadette McCrory for her patience and faith in me. I would like to thank my current committee members: Dr. Faraz Dadgostari, Dr. Bradley Whitaker, and Dr. Elizabeth Johnson. I would also like to extend my gratitude to past and current faculty, particularly Dr. Sage Kittelman, Mrs. Staci Turoski, Dr. William Schell, Dr. Frank Marchak, and Dr. David Claudio. I am deeply indebted to the Office of International Program staff, especially Mrs. Deborah Chiolero, and the Athletics Department staff who provided me with the resources to succeed on my academic and athletic journey. I would like to thank Mr. Kyle Kohn and the staff of the Montana Hospital Association who enabled me to join their team and gain valuable experiences. My deepest gratitude to all my classmates, teammates, coaches, and friends for alleviating and cheering me on during the toughest time of my research. Last, but not least, I would like to thank my brother, Rok Sitar, for being my best friend and a role model throughout my academic journey.

TABLE OF CONTENTS

TABLE OF CONTENTS CONTINUED

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

GLOSSARY

ASA	American Society of Anesthesiologists
BC	Borda Count
BMI	Body Mass Index
CAH	Critical Access Hospital
CAN	Certified Nurse Assistant
CMS	Center for Medicare & Medicaid Services
COPD	Chronic Obstructive Pulmonary Disease
CP	Complexity Index
CPT	Current procedural terminology
DASI	Duke Activity Status Index
DOA	Day of Admission
DOS	Day of Surgery
DOSA	Day of Surgery Admission
DT	Decision Tree
EACH	Essential Access Community Hospital
FMEA	Failure Mode Effect Analysis
HT	Hypertension
IRB	Institutional Review Board
LOS	Length of Stay
MAE	Mean Square Error
MAPE	Mean Absolute Error

GLOSSARY CONTINUED

MCDM	Multi-Criteria Decision Making
MDH	Medicare Dependent Hospital
ML	Machine Learning
NSQIP	National Surgical Quality Improvement Program
PSH	Prospective Payment System Hospitals
RF	Random Forest
RMSE	Root Mean Square Error
RRC	Rural Referral Centers
RUCA	Rural Urban Community Code
SCH	Sole Community Hospital
SRC	Surgical Risk Calculator
SSI	Surgical Site Infection
THR	Total Hip Replacement
TJ	Total Joint
TJA	Total Joint Arthroplasty
TKR	Total Knee Replacement
UA	Urban Areas
UI	User Interface
VA	Veterans Affairs
VTE	Venous Thromboembolism

ABSTRACT

Rural healthcare is represented by approximately one-third of community hospitals in the United States primarily in the Midwest and Western United States. Due to the lack of resources and the demographic characteristics of rural populations, rural community hospitals are under constant pressure to meet Center for Medicare & Medicaid Services (CMS) quality requirements. Meeting CMS quality requirements is particularly challenging in surgical care, due to the lower volumes and research opportunities, in addition to a shortage of qualified surgical specialists. The perioperative surgical home (PSH) model was established as a health management concept in a rural community hospital located in the Northwest of the United States to improve the quality of care by providing a longitudinal approach to patient treatment. The main opportunities for PSH improvement were identified in the "decision for surgery," "preoperative," and "postoperative" stages of the PSH model.

To improve PSH clinic performance this thesis proposes an improved National Surgical Quality Improvement Program (NSQIP) calculator User Interface (UI), as well as a new prediction model for predicting total joint arthroplasty (TJA) Length of Stay (LOS). The improved layout of the NSQIP calculator was developed based on two approved surveys by card sorting and Borda count methodology, while the new prediction model for predicting TJA patients' LOS was based on the Decision Tree (DT) machine learning model. A usability study of the NSQIP calculator UI identified opportunities for future improvements, such as the reorganized layout of postoperative complications and the addition of a supporting tool that would clearly define postoperative complications. The new DT prediction model outperformed a currently used NSQIP calculator in the prediction accuracy of TJA LOS, as it resulted in lower Root-mean-Square-Error values. Furthermore, the structure of the DT model allowed better interpretability of the decision-making process compared to the NSQIP calculator, which increased the trust and reliability of the calculated prediction. Despite some limitations such as a small sample size, this study provided valuable information for future improvements in rural healthcare, that would enable Rural Community Hospitals to better predict future outcomes and meet the strict CMS quality standard.

CHAPTER ONE

INTRODUCTION TO RURAL HEALTHCARE

Specifics of Rural Healthcare

Rurality is an abstract concept commonly associated with rolling hills and farmlands by the average American. However, the Census Bureau defines rural areas as any population, residents, or territory not in an urban area [1]. Urbanized Areas (UAs) are defined as any area of 50,000 or more people, while Urban Clusters (UCs) are areas with at least 2,500 and less than 5,000 people [1]. An alternative way to define the rurality of an area is by Rural-Urban Commuting Area (RUCA) Codes. RUCA codes are a census tract-based classification that uses standard census measures of population density, levels of urbanization, and journey-to-work commuting to characterize all U.S. census tracts with respect to their rural/urban status and commuting relationship to other census tracts [2]. In RUCA, each tract is assigned a primary code depending on the rural-urban characteristics (Table 1) [2].

Table 1. Rural-Urban Commuting Area (RUCA) code classification [2]

RUCA Code	**Classification Description (split the table)**
1	Metropolitan area core: primary flow within an urbanized area (UA)
2	Metropolitan area high commuting: primary flow 30% or more UA
3	Metropolitan area low commuting: primary flow 10% to 30% to a UA
4	Micropolitan area core: primary flow 30% or more to a large UC
5	Micropolitan high commuting: primary flow 30% or more to a large Urban Cluster (UC)
6	Micropolitan low commuting: primary flow 10% to 30% to a large UC

Table 1 Continued

RUCA Code	**Classification Description (split the table)**
7	Small town core: primary flow within an urban cluster of 2,500 to 9,999 (small UC)
8	Small town high commuting: primary flow 30% or more to a small UC
9	Small town low commuting: primary flow 10% to 30% to a small UC
10	Rural areas: primary flow to a tract outside a UA or UC
00	Not code: Census tract has zero population and no rural-urban identifies information

According to the classification of rural counties based on their major economic base, the following categories of rural areas have been established:

1) Agricultural counties – highly remote rural counties concentrated in the Northwest central census region;

2) Manufacturing counties – rural counties mainly located in the Southeast and near metropolitan areas;

3) Mining, oil, and energy counties – rural counties, such as coal-producing parts of Appalachia and oil-producing parts of northern mountain regions, that are fairly remote from metropolitan areas;

4) Counties specializing in governmental functions – rural counties spread throughout the country and generally more urbanized;

5) Persistent poverty counties – rural counties concentrated in the Southeast and considered highly rural areas with sparse populations;

6) Federal lands counties – rural counties, mainly in the West, with small, remote towns and cities;

7) Retirement community counties – rural counties in remote areas in California, Florida, Michigan, and the Southwest [3].

The primary focus of this thesis is a rural community hospital (RCH) in the Northwest. Community hospitals are all nonfederal, short-term general, and other special hospitals whose facilities and services are available to the public [3]. Approximately one-third of the community hospitals in the United States serve rural communities (Figure 1). Rural hospitals have an average capacity of 86 beds, which is roughly one-third the capacity of urban hospitals with 252 beds [4]. Furthermore, almost half of the RCH are facilities with no more than 25 beds (Figure 1) [4].

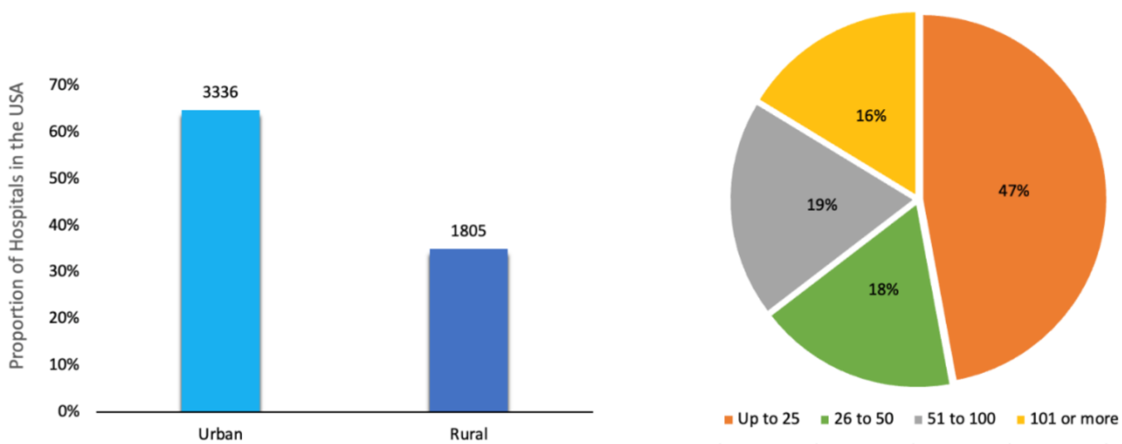


Figure 1. Proportion of Community Hospitals in Urban and Rural areas (left); Classification of Rural Community Hospitals based on bed number (right)

RCHs play a crucial role in the American healthcare system as it provides healthcare access for Americans living in rural areas. The importance of rural healthcare facilities was recognized by federal policymakers in the 1980s and 1990s, resulting in five special payment designations under Medicare: Critical Access Hospitals (CAHs), Medicare Dependent Hospitals (MDHs), Sole Community Hospitals (SCHs), Essential Access Community Hospitals (EACH), and Rural

Referral Centers (RRC) [5]. Rural hospitals provide approximately one-fifth of the major inpatient services received by urban community hospitals [3]. Furthermore, rural community hospitals provide the following services to patients: 1) General medical services; 2) Emergency medical services; 3) Surgical services 4) Pediatric services; and 5) Obstetric and perinatal services [3]. Hospitals can be further classified based on the resources available in trauma centers into five Trauma Levels. The Trauma Center designation is a process outlined and developed at a state or local level based on the unique criteria defined by selected state municipalities [6]. A Level I Trauma Center is a comprehensive regional resource that is capable of providing total care for every aspect of injury, from prevention through rehabilitation, while a Level V Trauma Center provides only initial evaluation, stabilization, and diagnostic medical services [6]. In certain states with low population density, rural hospitals offer only Level III – Level V trauma facilities. This means that if residents need Level I Trauma Care, they would need to leave the state to receive adequate treatment [7].

There is a limited amount of research conducted on surgical services provided by rural hospitals due to a lack of resources, qualified personnel, research centers, and funding [3]. Additionally, the lack of academic research involvement in rural hospitals was a common reason for physicians to select careers in an urban environment, where they have more opportunities to hone their specialized skills [8]. The greatest barrier of rural community hospitals is not in materialistic resources (i.e., facilities, equipment. etc.), but in the lack of surgeons and other supporting staff necessary for surgical procedures [3]. As a result, rural community hospitals are often challenged to achieve the same level of quality and safety as urban centers [3]. Providing a full range of quality medical services cost-effectively is a challenge that many RCHs face regularly. Furthermore, since RCHs have a larger proportion of Medicare discharges per total admissions than urban hospitals [3], RCHs must strictly follow the Medicare quality standards [9]. Risk prevention (i.e., hospital readmission, infections, etc.) is one of the main components of quality of care highly emphasized by the Centers for Medicaid & Medicare Services (CMS) [10]. Because healthcare risk characteristics include complexity and potential harm to patients, it is important to proactively manage healthcare failures [11]. To prevent healthcare-associated risks from occurring, rural community hospitals use several strategies to implement effective quality assurance programs, which include a consolidated quality assurance committee, collaborative arrangements with other hospitals, and risk management for quality assurance mechanisms [3].

Thesis Framework

Chapter Two will review the literature on the implementation of process improvement methodologies (Lean, Six Sigma, and Perioperative Surgical Home) in rural healthcare, define methodological tools and techniques used in this thesis, and introduce the Perioperative Surgical Home. Chapter Three will address the barriers of the NSQIP surgical calculator user interface, which is a frequently used tool in the Perioperative Surgical Home. Chapter Four will propose a new decision tree model as an alternative to the NSQIP calculator for predicting a patient's length of stay in a rural community hospital. Finally, Chapter Five will discuss the thesis outcomes, limitations, and future opportunities.

Thesis Goal and Objectives

The main goal of this thesis is to improve the services of a rural perioperative surgical home “Decision for Surgery,” “Preoperative,” and “Postoperative” stages by addressing the usability barriers of the NSQIP risk calculator and proposing a new prediction model for hospital length of stay after total arthroplasty (Figure 2)

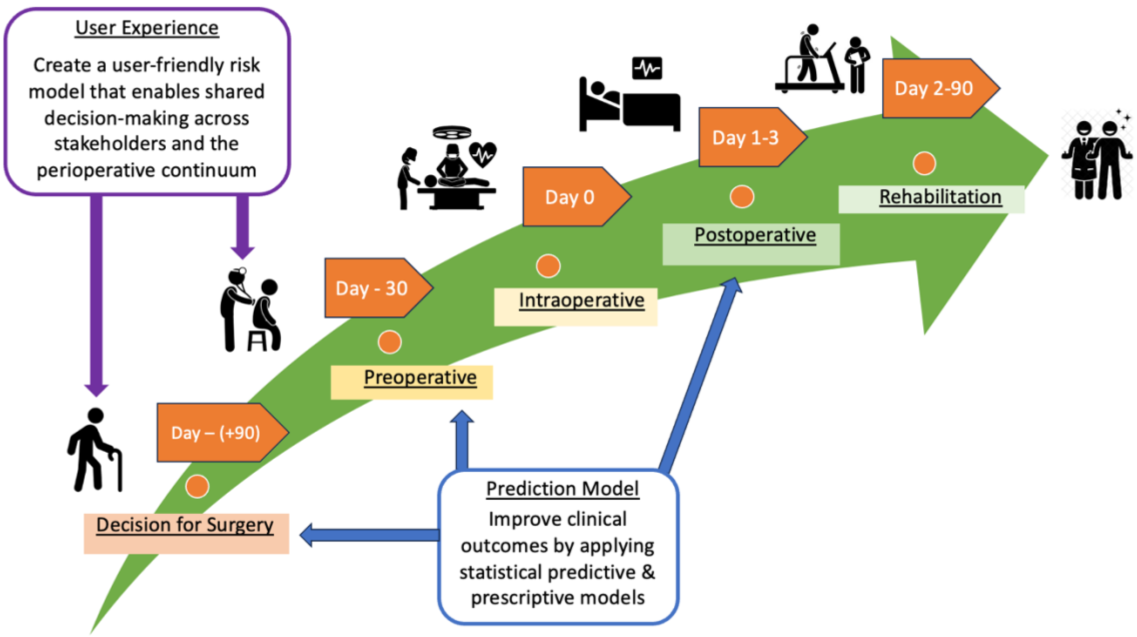


Figure 2. Conceptual model of Thesis framework

Objective 1: Analyze the usability of the NSQIP calculator's user interface and users' mental models of the displayed content.

This objective aimed to identify the optimal layout that matched users' mental models of the content displayed on the NSQIP calculator user interface.

Hypothesis 1: It was hypothesized that the ranking of postoperative complications displayed on the NSQIP calculator user interface would be inconsistent across different user types.

Expected Outcome 1: It was expected that expert and non-expert participants would inconsistently group/rank postsurgical complications on an NSQIP calculator user interface leading to a lack evidence-based of shared decision-making.

*Hypothesis 2***:** It was hypothesized that there is a lack of understanding of postsurgical complications outputted by the NSQIP calculator among non-expert participants.

Expected Outcome 2: It was expected that a high deviation of ranking between expert and non-expert participants would be observed among survey responses decreasing the usability of the NSQIP calculator.

Objective 2: Compare the prediction accuracy of Total Joint Arthroplasty (TJA) Length of Stay (LOS) between a custom decision tree model and NSQIP calculator for rural patients. The purpose of this objective is to demonstrate how machine learning (i.e., custom DT) can produce more accurate predictions with enhanced interpretability compared to the NSQIP calculator.

Hypothesis 3: It was hypothesized that a decision tree (DT) prediction model would result in a more accurate prediction of LOS compared to an NSQIP calculator.

Expected Outcome 3: It was expected that the DT model prediction would result in a lower Root Mean Square Error (RMSE) value compared to the NSQIP leading to better surgical optimization across the perioperative timespan.

CHAPTER TWO

PROCESS IMPROVEMENT METHODOLOGIES IN HEALTHCARE

Implication of Lean and Six Sigma on the Surgical Environment

Finding a balance between safe, reliable, and competent care while simultaneously improving the efficiency of processes and minimizing required resources is an ongoing challenge for many healthcare organizations [12]. Operating room and surgical procedures account for a large proportion of hospital operational expenses, providing many opportunities for system improvements [13]. To address the system barriers, healthcare organizations advocate process improvement methodologies to increase value, eliminate waste, and reduce costs as other industries have done [14]. Two prominent methodologies are Lean and Six Sigma, which are often combined into Lean Six Sigma [14].

The Lean Methodology

The Lean methodology was derived from the Toyota Production System, which enhances manufacturing performance by delivering quality products and services and reducing waste [15]. Lean focuses on improving systematic customer value delivery in conjunction with optimizing system performance [16]. Lean methodology classifies activities as value-added (VA) and non-value-added (NVA) [15]. The VA activities are a product or service for which customers are willing to pay, whereas NVA activities are those for which a customer would not be willing to pay [16]. To reduce systematic inefficiencies in healthcare systems (i.e., increase VA activities, and reduce NVA activities) Ohno [15] proposed the following seven wastes that were adapted to meet healthcare needs: overproduction (e.g., re-coding the same information multiple times and ordering unnecessary investigations); waiting (for patients, theatre staff, results, prescriptions and medicines, and discharge); transportation (patients, specimens and materials); overburden (stress, overworked staff); inventory (excess stock); motion (unnecessary staff movement looking for paperwork, supplies or people); and defects (readmission, repeated tests, and medical errors) [16].

The Six Sigma Methodology

Six Sigma is a quality improvement statistical technique that originated at Motorola [17]. The term sigma (σ) in the name Six Sigma, is a Greek letter used to describe variability, which represents an indicator of how often the defects are likely to occur [18]. Six Sigma is a numerical parameter used to identify process barriers, allowing process improvement teams to prioritize their work and focus on the projects with the greatest opportunities [18]. Six Sigma engages senior leaders in healthcare organizations and leverages dedicated resources for quality improvement projects often targeted to patient care and financial performance [19].

System improvement methods like Lean and Six Sigma were associated with the following strengths in complex healthcare systems [20]. First, Lean and Six Sigma objectively set out the processes where investigators need to understand the problem and guide decision-making; second, their processes are flexible and can be focused on improving a wide range of quality measures; third, both methodologies provide a framework to identify potential problem areas; finally, Lean and Six Sigma promote continual assessment and reassessment that allows adjustability and flexibility on the changes of needs and resources longitudinally [20]. It is a common practice that Lean and Six Sigma are used simultaneously in process improvement work. Where Lean methodology uses a reiterative cycle of improvements focused on mapping out the process and identifying VA and NVA activities, Six Sigma parameters complement the Lean methodology by identifying process steps with high levels of variability (i.e., the greatest opportunities for process improvement) [20].

Lean and Six Sigma methodologies were implemented at multiple steps throughout the research process. The main objective of the NSQIP calculator UI study was to reduce the variability of the user's mental model while using the NSQIP calculator. The desired outcome of the study was to develop a standardized layout of postoperative complications that would reduce non-value-added energy and time spent by the NSQIP calculator users, trying to identify the severity of postoperative complications. Furthermore, the lean principles were utilized in the methodological procedure of the NSQIP calculator usability study. Before extensive data collection, a pilot survey was tested with five participants. The pilot survey resulted in more than 15 different grouping categories which would present a barrier in the data-analysis procedure. To standardize the data-gathering process, the variability of potential grouping categories was reduced by manually pre-selecting four severity categories.

Lean principles were implemented during the TJA patient's LOS prediction model development. The objective of this study was not only to identify a prediction model that would adequately predict patients' LOS but also to identify prediction variables with a significant impact on LOS. By identifying variables with a large impact on hospital LOS (value-added variables), clinicians will be able to better allocate their resources in the preoperative stage of the PSH model, and consequently impact patient LOS. Six-sigma methodology principles were implemented in evaluating the prediction accuracy of the decision tree model and NSQIP calculator. The desired outcome of any regression prediction model is to minimize the Root-Mean-Square-Error (RMSE). The RMSE is a key performance indicator that describes the variability (error/deviation) from the mean value. Aligned with the six-sigma methodology, minimization of the numerical variability parameter in a process or prediction is a key to successful and sustainable outcomes.

Perioperative Surgical Home

Increasing costs combined with non-optimal healthcare outcomes led the Centers for Medicare & Medicaid Services (CMS) to reconstruct payment methods. Specifically, 80% of all CMS payments depend on value parameters, with the overall goal to mandate hospitals, groups, and providers to decrease their expenses, and improve patient experience and clinical outcomes [21]. In surgical care, a new delivery care model referred to as the Perioperative Surgical Home (PSH) was developed by the American Society of Anesthesiologists (ASA) as an addition to the Lean and Six Sigma methodologies to optimize operative services and procedural outcomes [21]. One of the most frequently used definitions of PSH was developed by the University of Alabama at Birmingham, which described PSH as “An innovative, patient-centered, surgical continuity of care model that incorporates shared decision-making” [22]. The overall goal of PSH is to target the Institute for Healthcare Improvement (IHI) “Triple Aim” (Figure 3) [23].

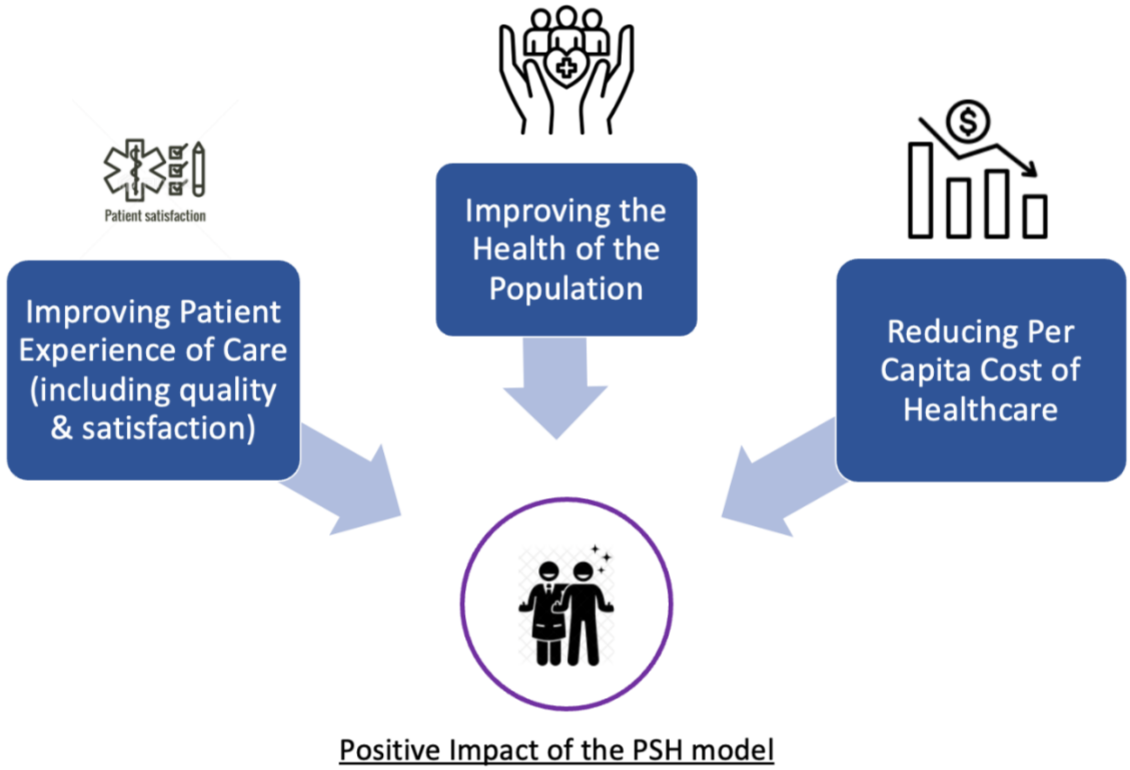


Figure 3. Institute for Healthcare Improvement "Triple Aim"

Conceptually, the PSH model aims to reduce the variability in perioperative care by having one team led by an anesthesiologist communicate with other physicians in all aspects of surgical care [24]. Anesthesiologists, who are uniquely integrated to manage complex care, have demonstrated the ability to serve while coordinating other leadership functions including resource allocation, staffing, scheduling, and accounting for patient care [21]. Coordinated care through physician co-management is accomplished through effective communication, shared decision-making, and standardized PAS processes [21]. Compared to a traditional surgical approach, preoperative, intraoperative, postoperative, and post-discharge periods are treated as single episodes. In this way, PSH assures a continuum of care rather than discrete episodes [24]. Postsurgical authorization is made, and patients enter the five-phase continuum process of PSH (Figure 4) [24].

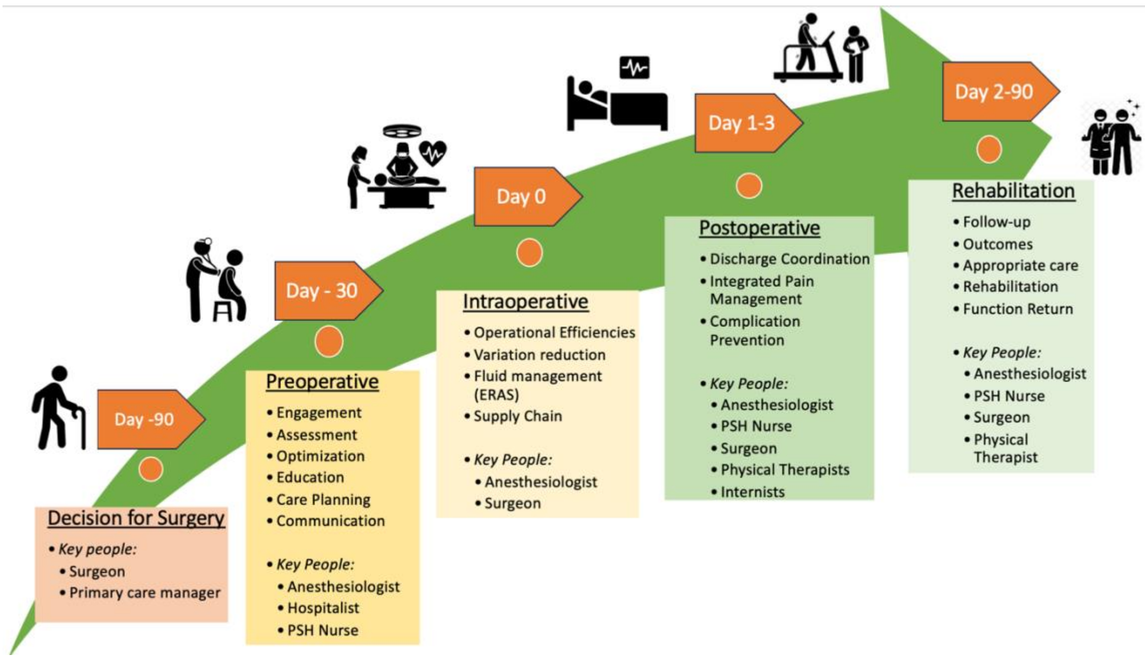


Figure 4 Perioperative Surgical Home Model [24]

The PSH clinic was implemented in a rural community hospital, a licensed Level-III trauma center, that provides services primarily to three rural counties. The hospital service area spans an area of 9,000 square miles and a population of 134,357 residents [25]. The PSH clinic was successfully initiated, overcoming a lack of resources due to a strong team-based approach among clinicians. The PSH clinic's multidisciplinary team consisted of an anesthesiologist, surgeon, registered nurse, physician assistant, hospitalist, and primary care physician. The communication within the team was performed using in-person meetings and clinical notes. Frequent and regular communication is a foundation for the co-management model used in the

PSH clinic, which promotes joint responsibility to manage the healthcare of an individual patient [26].

Total Joint Arthroplasty

Total joint arthroplasties (TJA) are the most common elective surgical procedures in the U.S. [27]. According to Stanford Medicine, the three most common types of TJA surgeries are hip replacement, knee replacement, and shoulder replacement [28]. With the aging of the “baby boomers,” higher rates of diagnosis, treatment of advanced arthritis, and growing demand for improved mobility and quality of life, the annual volumes of TJA surgeries are expected to significantly increase [29]. The prevalence of total hip and knee replacement in the US Northwest was greater than 1.20% for total hip replacement (THR), and between 1.5% - 1.75% for total knee replacement (TKR) surgeries [29]. To facilitate this large demand, a PSH in Northwest is utilizing the NSQIP prediction model, which enables better utilization of available resources and consequently a better quality of care for TJA patients [30].

Introduction of the NSQIP Calculator

The NSQIP was launched by the Department of Veterans Affairs (VA) in 1994 and has since expanded into the private sector through the efforts of the American College of Surgeons (ACS) [31]. The NSQIP is the first nationally validated, outcome-based, risk-adjusted, peer-controlled program for the measurement and enhancement of quality surgical care [31]. The NSQIP database includes risk factors and outcome information on approximately 100,000 major surgical procedures annually [31]. The overall purpose of the NSQIP database is to assess surgical quality at more than 200 hospitals [32]. Furthermore, the database serves as a resource for surgeons and researchers who want to research surgery-related topics [31].

The ACS used the NSQIP database to create a risk-predicting calculator [33]. The patient cohorts supplying the database were collected from a teaching or academic affiliation and large hospitals with a bed size exceeding 500 beds [34]. The NSQIP calculator is a tool that provides accurate, patient-specific risk information to guide both surgical decision-making and informed consent [35]. The NSQIP risk calculator is based on a logistic regression model that was developed based on 1.4 million operative procedures performed in approximately 500 hospitals in the US [32]. Logistic regression estimates the probability of an event occurring based on the given dataset of independent variables, where the variable is bounded between 0 and 1. The strongest NSQIP calculator predictor is a current procedural terminology (CPT) code, which allows the surgical risk calculator to effectively make predictions for a wide range of operations [36]. The CPT code risk adjusts for the operative procedure in the performance assessment (i.e., open-heart surgery will result in greater risks compared to the Laparoscopy procedure) [37]. The risk calculator also uses 20 patient predictors (Table 2) with the CPT code to predict the likelihood of 18 (Table 3) patient outcomes occurring within 30 days of surgery [35]. Additionally, the NSQIP database is constantly updating, which directly impacts the NSQIP calculator outputs [38].

Table 2. NSQIP Calculator Inputs [35]

	Independent Variable	**Variable description**
NSQIP Variables	Specific CPT class (Specific Procedural Terminology) Class	Current Procedural Terminology (CPT) code is a medical terminology used to describe the type of medical procedure[39]
	Age	Patient age in years
	Gender	Patient gender (male, female)

Table 2 Continued

	Independent Variable	Variable Description
NSQIP Variables	Functional status [35]	Functional status/level of self-care demonstrated by the patient within the 30 days prior surgery • Independent: The patient does not require assistance from another person for any activities of daily living. This includes a person who is able to function independently with prosthetics, equipment, or devices • Partially dependent: The patient requires some assistance from another person for activities of daily living. • Totally dependent: The patient requires total assistance for all activities of daily living.
	Emergency Case [35]	The principal operative procedure must be performed during the hospital admission for diagnosis AND the surgeon and/or anesthesiologist must report the case as emergent
	Steroid use for chronic conditions [35]	Regular administration of oral or parenteral corticosteroid medication or immunosuppressants for a chronic medical condition, within the 30 days prior to surgery, or at the time the patient is being considered as a candidate for surgery. A one-time pulse limited short course or a taper of fewer than 10 days duration would not qualify. Long-interval injections of long-acting agents would qualify.
	Ascites within 30 days prior to surgery [35]	The presence of fluid accumulation in the peritoneal cavity was noted on physical examination, abdominal ultrasound, or abdominal CT/MRI within 30 days prior to surgery. Documentation must state either active or history of liver disease or must state secondary to malignancy.
	Systematic sepsis within 48 hours prior to surgery [35]	Any of the following occurring within 48 hours prior to surgery: • Systematic Inflammatory Response Syndrome (SIRS) • Sepsis • Septic Shock
	Ventilator dependent	A patient requiring ventilator-assisted respiration at any time during the 48 hours preceding surgery. This does not include the treatment of sleep apnea with CPAP

Table 2 Continued

	Independent Variable	**Variable Description**
NSQIP Variables	Disseminated cancer	The patient has primary cancer that has metastasized to a major organ AND meets at least one of the following: • Active treatment for cancer within one year of the surgery date. If the surgical procedure is the treatment for metastatic cancer, answer "Yes". • The patient has elected not to receive treatment for the metastatic disease. • The patient's metastatic cancer has been deemed untreatable. Report the following cancers as Disseminated Cancer: Acute Lymphocytic leukemia (ALL), Acute Myelogenous Leukemia (AML), and Stage IV Lymphoma. Do not report the following as Disseminated Cancer: Chronic Lymphocytic Leukemia (CLL), Chronic Myelogenous Leukemia (CML), Stage I through III Lymphomas, or Multiple Myeloma
	Congestive heart failure in 30 days prior to surgery	Only newly diagnosed CHF within the previous 30 days or a diagnosis of chronic CHF with signs or symptoms of CHF in the 30 days prior to surgery fulfills this definition
	History of severe COPD	Chronic obstructive pulmonary disease (COPD) (such as emphysema and/or chronic bronchitis) resulting in one or more of the following: • Functional disability from COPD (for example, dyspnea, inability to perform ADLs) • Hospitalization in the past for treatment of COPD • Chronic bronchodilator therapy with oral or inhaled agents • FEV1 of <75% of predicted • Do not include patients whose only pulmonary disease is asthma • Do not include patients with diffuse interstitial fibrosis or sarcoidosis

Table 2 Continued

	Independent variable	**Deascription**
NSQIP Variables	Dialysis	Acute or chronic renal failure requiring treatment with peritoneal dialysis, hemodialysis, hemofiltration, hemodiafiltration, or ultrafiltration within 2 weeks prior to surgery. If a patient requires dialysis but refuses it, the answer to this variable will be "Yes"
	Acute renal failure	A clinical condition associated with the rapid decline of kidney function. The patient meets one of the following: • Increased BUN on two measurements AND two Cr results > 3mg/dl • The surgeon or physician has documented Acute Renal Failure AND one of the following: ○ Increased BUN on two measurements ○ Two Cr results > 3mg/dl
	ASA Class	• ASA 1: Normal healthy patient. • ASA 2: Patient with the mild systematic disease. • ASA 3: Patient with the severe systematic disease. • ASA 4: Patient with a severe systematic disease that is a constant threat to life • ASA 5: Moribund patient who is not expected to survive without the operation
	Diabetes status	The individual requires daily dosages of exogenous parenteral insulin or an oral hypoglycemic agent to prevent hyperglycemia. A patient is not included if diabetes is controlled by diet alone.
	HT Medicine (HT - Hypertension status)	The patient has a diagnosis of HTN in the medical record and will require antihypertensive medication(s) within 30 days prior to surgery
	Dyspnea	The patient's dyspnea status when they were in their usual state of health, prior to the onset of the acute illness, within the 30 days prior to the time the patient is being considered a candidate for surgery

Table 2 Continued

	Independent variable	**Description**
NSQIP Variables	Smoking status	The patient smoked cigarettes in the year prior to admission for surgery. Patients who smoke cigars or pipes or use chewing tobacco are not included
	BMI	A measure of weight adjusted for height (lb/ft) [40]
Additional variables	Distance	Distance traveled to the Rural Community Hospital (miles)
	Duke Activity Status Index (DASI)	Duke Activity Status Index (DASI) developed to measure patient overall functional capacity (points scale 0-52.8) [41]
	Education: TJ class (Total joint class)	Education: The Total Joint Replacement (TJR) class covers pre- and postsurgical information to educate patients and reduce his/her level of fear and concern before surgery [42]
	Household Status	Patient primary living status (lives alone, living with another) [43]
	Zip Code (primary)	Zip codes are spatially based, hierarchical codes that allow postal service to route and deliver mail [44]
	Insurance	A policy that pays for patients' treatment expanses [45]
	Income	Patients' household income (annual income in $) [30]
	Pre-operative duration	The duration between first preoperative visit to the hospital and the day of the surgery (days)

Table 3. NSQIP Calculator Outputs

Postoperative Complication	**Definition**
Serious Complication	Cardiac arrest, myocardial infarction, pneumonia, progressive renal insufficiency, acute renal failure, PE, DVT, return to the operating room, deep incisional SSI, organ space SSI, systematic sepsis, unplanned intubation, UITI, wound disruption [35]
Any Complication	Superficial incisional SSI, deep incisional SSI, organ space SSI, wound disruption, pneumonia, unplanned intubation, PE, DVT, ventilator > 48 hours, progressive renal insufficiency, acute renal failure, UTI, stroke, cardiac arrest, myocardial infarction, return to the operating room, systemic sepsis [35]
Pneumonia	Includes cardiac arrest or myocardial infarction [35]

Table 3 Continued

Postoperative Complication	**Definition**
	• Cardiac arrest: The absence of cardiac rhythm or the presence of a chaotic cardiac rhythm requiring the initiation of CPR, which includes chest compressions • Myocardial infarction: ECG changes, new elevation in troponin, or physician diagnosis
Cardiac Complication	Cardiac complications are defined as cardiac death, pump failure (Killip grade ≥ II), sustained ventricular tachycardia on fibrillation (SVT/VF), and advanced atrioventricular block (AVB) [46]
Surgical Site Infection (SSI)	Includes superficial incisional SSI, deep incisional SSI, or organ space SSI [35] • Superficial Incisional SSI: Infection that involves only skin or subcutaneous tissue of the incision. It also includes either: purulent drainage, positive culture, signs/symptoms of infection, and an incision that is deliberately opened by the surgeon or diagnosis by the attending physician. • Deep Incisional SSI: Infection that appears to be related to the operation and involves deep soft tissues (for example, fascial and muscle layers) of the incision. It also includes either purulent drainage, spontaneous dehiscence, deliberate opening by the surgeon, abscess involving the deep incision, or diagnosis by the attending physician. • Organ Space SSI: Infection that involves any part of the anatomy (for example, organs or spaces), other than the incision, which was opened or manipulated during an operation. It also includes either: purulent drainage, positive culture, absence, or diagnosis by the attending physician
Urinary Tract Infection	Bladder infection is diagnosed using a combination of clinical symptoms and laboratory confirmation (e.g., urine culture, pyuria, positive dipstick) or the initiation of appropriate antimicrobial therapy [35]
Venous Thromboembolism	The identification of a new thrombus within the venous system is described in studies as present in the superficial or deep venous systems but requires therapy. This diagnosis is confirmed by duplex, venogram, CT scan, or other imaging modality, AND the patient requires treatment with anticoagulation therapy and/or placement of a vena cave filter or clipping of the vena cava [35]

Table 3 Continued

Postoperative Complication	**Definition**
Renal Failure	Includes either progressive renal insufficiency OR acute renal failure requiring dialysis [35] • Progressive renal insufficiency: a rise in creatinine of > 2 mg/dl from preoperative value, but with no requirement for dialysis. Acute renal failure requiring dialysis: A patient who did not require dialysis preoperatively, worsening of renal dysfunction postoperatively requiring hemodialysis, peritoneal dialysis, hemofiltration, hemodiafiltration, or ultrafiltration.
Ileus	• Prolonged Postoperative NPO or NGT Use: Prolonged NPO status or NGT use for suctioning or decompression, more than 3 days postop (POD4, or later) OR reinsertion of NGT or reinstating NPO status any time POD4 or later within 30 days [35]
Anastomotic Leak	There was a leak of endoluminal contents through an anastomosis. This could include air, fluid, GI contents, or contrast material. The presence of an infection/abscess thought to be related to an anastomosis, even if the leak cannot be definitively identified as visualized during an operation, or by contrast, extravasation would still be considered an anastomotic leak if this is indicated by the surgeon [35]
Readmission	Readmission rate as a marker of the quality of hospital care it is defined as "the number of patients who experienced unintended, acute readmission or death within 30-days of discharge from the index admission, divided by the total number of patients discharged alive within the reference period' [47]
Return to OR	Return to the operating room for additional surgery that was not planned at the time of the initial surgery [35]
Death	Human death is the irreversible loss of functioning of the organism as a whole [48]
Discharge to Nursing or Rehab Facility	Includes discharge to one of the following facilities [35]: • A skilled care facility that was not home previously (sub-acute hospital, skilled nursing home/facility, transitional care unit, long term care facility, or ventilator bed • An unskilled care facility that was not home previously (unskilled nursing home or assisted facility) • Rehab (inpatient rehabilitation facility including rehabilitation distinct part units of a hospital) Separate acute care facility

Table 3 Continued

Postoperative Complication	**Definition**
Sepsis	Is the systematic inflammatory response to infection [49]
Length of Stay (LOS)	The LOS is defined as the number of days an in-patient will remain in hospital [50]

User Interface

The user interface (UI) is the point of human-computer interaction and communication allowing users to work with machines in a diverse set of industries [51]. The role of UI architects is to develop processes, organization, and artifacts that facilitate users' mental models and interaction with the product (i.e., NSQIP calculator) without constraining the product for a particular group of users (i.e., novice users, colorblind, etc.) [52]. The quality of the UI is most frequently evaluated with 10 usability guidelines (i.e., heuristic assessment) developed by Jakob Nielsen: 1) Visibility of System; 2) Match Between System and the Real World; 3) User Control and Freedom; 4) Consistency and Standards; 5) Error Prevention; 6) Recognition Rather than Recall; 7) Flexibility, and Efficiency of Use; 8) Aesthetic and Minimalistic Design; 9) Help User Recognize, Diagnose, and Recover from Errors; and 10) Help and Document [53]. The advantage of Jacob Nielsen's heuristic evaluation is the small number of evaluators required to evaluate UI [53]. Contrary, the disadvantage of this method is that all UI evaluators need to be highly educated and experienced in their area of expertise to successfully evaluate the UI [53].

An alternative approach to evaluate UIs is to capture feedback from product users. A frequently used technique to capture and organize users' feedback on the UI is card sorting [54]. As the name implies, the card-sorting method consists of concepts (cards), that participants sort into piles based on their similarities. After the cards are sorted into piles, the groups are labeled with the name that best describes the commonalities of the concepts in the group. In open card sorting, there is no restriction on the number and type of categories, while in closed card sorting there already exists pre-selected categories [55]. Due to the simplicity of use, focus on the subject's terminology (rather than that of external experts), and ability to capture tacit knowledge a card sorting methodology was selected as a primary data-gathering and grouping technique in this usability study [56]

The Borda Count (BC) methodology is a frequently used tool to rank and prioritize displayed content with a small number of participants (i.e., n<30) [57]. Let *i* represent the decision criteria and *j* the decision alternative. Criteria weights (w_i) and standardized criteria values (a_{ij}) are used to calculate a total weighted score for each decision alternative as follows (Eq. 1):

$$C_j = \sum_{i=1}^{I} w_i a_{ij} \qquad \text{Eq. 1}$$

The a_{ij} represents each ij^{th} combination, while w_i indicates a weight (power) associated with i^{th} decision criteria (i.e., Lowest ranked category will have a w_i=1; the highest ranked category will have a w_i=n) [56]. In other words, the most desirable alternative is multiplied by the total number of alternatives (n), while the least desirable alternative is multiplied by 1 [57]. The alternative with the highest score is ranked as number one, while the alternative with the lowest score is ranked as last [57]. Due to the multiple decision criteria, and multiple decision makers (i.e., study participants with different levels of expertise) involved in this study, the BC method was selected as an optimal tool to rank participants' responses [56].

Machine Learning

Interpretability of the NSQIP calculator user interface is not the only limitation of the calculator when used in rural healthcare. For the past two decades, big data (i.e., a massive amount of information) has become a topic of special interest in the healthcare industry [58]. Various sources of big data including hospital records, medical records of patients, and results of medical examinations, are all used to improve medical services and increase the quality of care [58]. Big data is frequently used for developing prediction models, which serve as an important resource at various stages of hospital care [30]. On the flip side, there are various barriers associated with handling big data which could only be surpassed with high-end computing algorithms for analyzing big data [58]. The NSQIP calculator predictions are based on a logistic regression model, which has proven to be less optimal for predicting healthcare outcomes compared to modern machine learning (ML) models [30]. As the name implies, ML models learn rules from the data, by sifting through a vast number of variables, looking for a combination that reliably predicts outcomes [59]. An advantage of ML models compared to regression models is the capability to handle enormous numbers of predictors and combine them in nonlinear interactive ways, which enables analyzing complex datasets [59].

Electronic data collection of medical records allows healthcare facilities to apply ML models to develop prediction of clinical outcomes. ML models require an adequate amount of data to calculate robust accurate prediction models [59]. As a result, ongoing research primarily focuses on applying ML models in urban healthcare facilities which possess more resources compared to rural hospitals [30]. Dr. Sridhar is one of the first researchers who implemented an ML model in a Rural clinical setting. A Random Forest (RF) model was used to predict TJA patient LOS. Study outcomes indicated that the RF model calculated a more accurate prediction of LOS compared to the NSQIP calculator [30]. Our work builds on the top of Dr. Sridhar, where we developed a custom-made decision tree model for predicting TJA patient LOS in the Rural Community Hospital.

A Decision Tree (DT) is a visual and analytical ML model, where each non-leaf node denotes a test on an attribute, and each branch indicates a result of a test [60]. Due to the interpretable and hierarchical structure of a DT model, it is one of the most used ML methodologies [60]. By testing the accuracy of a DT model prediction of a TJA patient's LOS, we expanded Dr. Sridhar's work, exploring if even the basic and highly interpretable ML model such as DT can still provide more accurate prediction outcomes compared to the NSQIP calculator.

Another specific component of this study was the application of the NSQIP calculator to rural healthcare. The NSQIP calculator is based on data collected from 200 major hospitals around the United States [32]. Patients living in urban areas with major hospital facilities (i.e., supplying the NSQIP database) significantly differ from the patients living in rural areas [61]. To address this barrier and improve patient care in rural healthcare that would better fit the needs of their residents, a new prediction model for predicting patient Length of Stay (LOS) was developed.

CHAPTER THREE

SYSTEMATIC RANKING OF NSQIP SURGICAL RISK CALCULATOR OUTPUTS

Contribution of Authors and Co-Authors

Author: Nejc Sitar

Contribution: Conceptualization, Formal Analysis, Methodology, Visualization, Writing

Co-Author: Bernadette McCrory

Contribution: Project Administration, Supervision, Validation, Review and Editing

Co-Author: Staci Turoski

Contribution: Supervision, Validation, Review and Editing

Co-Author: Caleb T. Gardner

Contribution: Data Collection

Co-Author: Corey Guildner

Contribution: Data Collection

Manuscript Information

Nejc Sitar, Staci Turoski, Bernadette McCrory, Caleb T. Gardner, Corey Guildner Jr.

Proceedings of the Human Factors and Ergonomics Society Annual Meeting

Status of Manuscript:
☐ Prepared for submission to a peer-reviewed journal
☐ Officially submitted to a peer-reviewed journal
x Accepted by a peer-reviewed journal
☐ Published in a peer-reviewed journal

Proceedings of the HFES Annual Meeting
8/7/2023
TBD
HFES 67th International Annual Meeting

Abstract

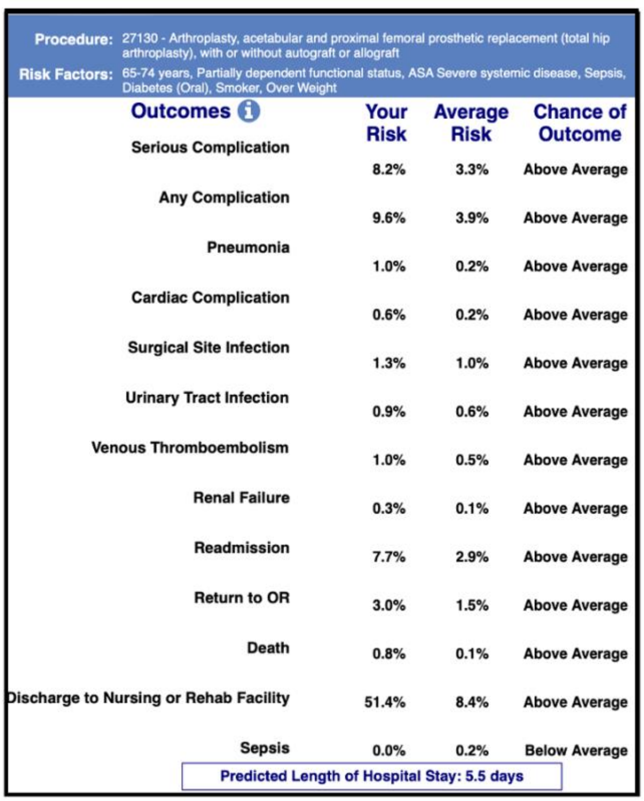
The National Surgical Quality Improvement Program (NSQIP) calculator is a risk-predicting calculator, that predicts the likelihood of postoperative complications. Card sorting methodology was used to capture feedback from expert and non-expert participants regarding the severity of surgical complications. The study result indicated ambiguity of postoperative complications. To overcome this barrier, a proposed order of postoperative complications was developed based on the Borda count ranking performed by surgeons.

Introduction

Individualized surgical risk prediction methods have become an important tool to support shared decision-making in surgical procedures [62]. Surgical risk-predicting calculators can predict the likelihood of risk for several postoperative complications based on the information entered including procedure types and individual patient factors, such as demographic characteristics and comorbidities [63]. One of the most frequently used surgical risk-predicting methods is the American College of Surgeons National Surgical Quality Improvement Program (NSQIP) Surgical Risk calculator, which became the gold standard for estimating the likelihood of postoperative risks [35]. The NSQIP calculator uses twenty different predictors, to calculate the likelihood of 18 different outcomes/complications within the 30-day postoperative period [35]. The calculator's predicted outcomes are displayed via a user interface (UI) in no specific order such as rank or risk severity [35]. For example, a male, 65-74 years of age with XXX and XX has XXX and YYY (Figure 5).

The UI is the point of human-computer interaction and communication particularly in complex and technical fields that interact with a diverse set of knowledge and novice users [51]. According to the Seeheim model [64], the UI consists of three functional blocks, “Presentation,” “Dialogue Control,” and “Application Interface”. The focus of this case study was to evaluate and understand the NSQIP’s presentation system because it could be challenging for someone with no medical background to prioritize the importance of the NSQIP calculator outputs.

A common technique used to capture and organize users' feedback on the presentation part of UI is the card sorting technique [54]. Card sorting is a data-gathering technique, that facilitates the systematic organization of content, features, and functions to discover an optimal organization of information on the UI [55].


Procedure: 27130 - Arthroplasty, acetabular and proximal femoral prosthetic replacement (total hip arthroplasty), with or without autograft or allograft
Risk Factors: 65-74 years, Partially dependent functional status, ASA Severe systemic disease, Sepsis, Diabetes (Oral), Smoker, Over Weight

Outcomes	Your Risk	Average Risk	Chance of Outcome
Serious Complication	8.2%	3.3%	Above Average
Any Complication	9.6%	3.9%	Above Average
Pneumonia	1.0%	0.2%	Above Average
Cardiac Complication	0.6%	0.2%	Above Average
Surgical Site Infection	1.3%	1.0%	Above Average
Urinary Tract Infection	0.9%	0.6%	Above Average
Venous Thromboembolism	1.0%	0.5%	Above Average
Renal Failure	0.3%	0.1%	Above Average
Readmission	7.7%	2.9%	Above Average
Return to OR	3.0%	1.5%	Above Average
Death	0.8%	0.1%	Above Average
Discharge to Nursing or Rehab Facility	51.4%	8.4%	Above Average
Sepsis	0.0%	0.2%	Below Average

Predicted Length of Hospital Stay: 5.5 days

Figure 5. NSQIP Calculator User Interface, Output

The decision-making method Borda Count (BC) is also a commonly used tool to rank and organize feedback for UI display [57]. The BC methodology is based on multiplying an “n” number of alternatives, where the most desirable alternative is multiplied by (n), the second most desired alternative is multiplied by (n-1), and the least desired alternative is multiplied by (1) [57]. Furthermore, BC was identified as one of the safest and optimal systems to conduct this ranking process compared with other ranking techniques [65].

This study was conducted in a rural healthcare environment to evaluate the usability of the UI components of the NSQIP calculator output. This study’s primary aim was to identify the ranking importance of postoperative complications displayed on the NSQIP calculator UI. A secondary aim was to identify the public/patient understanding of postoperative complications. It was hypothesized that there is a significant disparity in the organization and grouping of postoperative complications between different user groups of the NSQIP calculator.

Literature Review

Several studies addressed the problem of unclear understanding and lack of public education on some of the postoperative complications displayed by the NSQIP calculator [66]–[69]. Mackenzie argued that there is a lack of understanding of the definition and classification of pneumonia in public and scientific circles [66]. Ahmed et al. identified that 43.8% of study participants had a bad understanding of cardiac complications [67]. Based on the current literature, both urinary tract infection (UTI) and surgical site infection (SSI) were recognized as two complications that lack public understanding and awareness [70]. Furthermore, sepsis was recognized as the least familiar type of infection amongst adult Americans with ~88% of study participants never heard of the term Sepsis [68]. Finally, in a study conducted on evaluating global public awareness of Venous Thromboembolism (VTE), the authors identified that the general understanding of VTE in public was low [69]. Accordingly, this study focused on 16 postoperative complications displayed in the NSQIP output [35] (Table 4).

Table 4. Definition of postoperative complications

Postoperative complication	**Definition**
Serious Complication	Cardiac arrest, myocardial infarction, pneumonia, progressive renal insufficiency, acute renal failure, PE, DVT, return to the operating room, deep incisional SSI, organ space SSI, systematic sepsis, unplanned intubation, UTI, wound disruption.
Any Complication	Superficial incisional SSI, deep incisional SSI, organ space SSI, wound disruption, pneumonia, unplanned intubation, ventilator > 48 hours, progressive renal insufficiency, UTI, stroke.
Pneumonia	Includes cardiac arrest or myocardial infarction.
Cardiac Complication	Cardiac death, pump failure (Killip grade ≥ II), sustained ventricular tachycardia on fibrillation (SVT/VF), and advanced atrioventricular block [46]
Surgical Site Infection (SSI)	Includes superficial incisional SSI, deep incisional SSI, or organ space SSI.
Urinary Tract Infection	Bladder infection is diagnosed using a combination of clinical symptoms and laboratory confirmation or the initiation of appropriate antimicrobial therapy.
Venous Thromboembolism	Present in the superficial or deep venous systems but requires therapy. Confirmed by imaging modality. A patient requires treatment with anticoagulation therapy and/or placement of a vena cave filter or clipping of the vena cava.
Renal Failure	Includes either progressive renal insufficiency OR acute renal failure requiring dialysis
Ileus	Prolonged Postoperative NPO or NGT use.
Anastomotic Leak	A leak of endoluminal contents through an anastomosis. This could include air, fluid, GI contents, or contrast material.
Readmission	The number of patients who experienced unintended, acute readmission within 30-days of discharge [47]
Return to OR	Return to the operating room for additional surgery that was not planned at the time of the initial surgery.
Death	Human death is the irreversible loss of functioning of the organism.

Table 4 Continued

Postoperative complication	**Definition**
Discharge to Nursing or Rehab Facility	Includes discharge to one of the following facilities: A skilled care facility, nursing home/facility, an unskilled care facility, rehab Separate acute care facility
Sepsis	Systematic inflammatory response to infection [49]

The card sorting technique was frequently discussed among usability practitioners for software development [71], [72]. In his study, Fuccella, highlighted the ability of card sorting to identify a discrepancy in the website, software interface, or multimedia product, and to identify areas of potential improvements for existing and future users [71].

Nielsen added to the existing work of Fuccella, by proposing a second round of groupings into larger groups (i.e., group similar piles generated in the first stage into second-level headings, but not continue to the third-level grouping) [73]. Additionally, Nielsen did not recommend advanced statistical analysis for card sorting results but suggested data review to determine categories manually using a qualitative approach with up to 10 study participants [73].

Borda count (BC) was also a recurrent research topic in the last decade [74], [75]. Mula et. all used the BC methodology to identify and rank the most frequently used mobile functions that become a target for usability improvements in smartphone devices [74]. The authors concluded that the BC is an effective methodology when selecting alternatives requiring usability improvements, as it helped identify the priority and ranking of importance between alternatives [74].

Methods

Pilot Study

The open card sorting pilot survey was distributed to two experts (i.e., people who work in healthcare) and two non-expert participants (i.e., people who do not work in healthcare). Two specific NSQIP calculator outputs were provided to study participants, who were asked to group the postoperative complications based on their severity (i.e., open card sorting). The study resulted in diverse outcomes which indicated that more standardization needed to be implemented in the card sorting procedure that would allow further statistical analysis.

Main Study

Based on the pilot study outcomes, finalized surveys were developed to capture feedback from expert and non-expert study participants. Two separate surveys consisted of 5 demographic questions, one closed card-sorting question, and two content-specific questions. Study participants were asked to group 16 postoperative complications in the four pre-set card sorting categories developed in alignment with the Failure Mode and Effect Analysis (FMEA) severity classification [76]. The postoperative complications were classified into Catastrophic Severity, Major Severity, Moderate Severity, and Minor Severity groups, where the most severe postoperative complications were grouped in the Catastrophic Severity category [76]. Next, participants were asked to rank postoperative complications within each category in ascending order (i.e., Most severe complication = 1). Finally, the participants were asked to select any postoperative complications whose meaning is unclear to the public with no medical experience. The expert and not-expert surveys differed in two demographic and one survey-specific question. Both surveys were institutionally reviewed by the Institutional Review Board (IRB) [77] and developed in Qualtrics software [78], distributed to study participants by printed QR codes and emails.

Data Analysis

Data analysis was similar for expert and non-expert participants**,** except that non-expert participants were further classified into five categories based on the levels of familiarity with postoperative complications (Very High; High; Moderate; Low; and Very Low) [79]. The grouping of postoperative complications into four pre-set categories was completed by the level of participant familiarity and overall, where the severity category with the most votes was selected as a primary option. The BC ranking was performed on the survey results of all non-expert (i.e., no grouping by expertise) and expert participants.

Results

Pilot Study

A pilot study conducted on two experts, and two non-expert participants (open card sorting) resulted in 13 different severity categories. As a result of that, four pre-set categories were developed according to FMEA guidelines in an actual survey, to allow post-survey data analysis.

Participant Demographics

A non-expert survey was conducted on 30 (47%) male and 34 (53%) female participants (n=64). When asked about their level of familiarity with postoperative complications, four (6%) participants had a very high level of familiarity, five (8%) participants had a high level, 19 (30%) participants had a moderate level, 28 (44%) participants had a low level, and eight (13%)

participants had a very low level of familiarity. Additionally, 60 (94%) patients had never heard about the NSQIP calculator, and 57 (89%) participants were unfamiliar with at least one postoperative complication displayed on the NSQIP calculator user interface.

The expert participants consisted of one surgeon, one clinician, five nurses, a certified nursing assistant (CNA), a medical administrator, and a medicine student. Excluding the student, all experts had five or more years of professional experience. Seven expert participants were females (70%), and three expert participants were males (30%). Four experts (40%) were familiar with the NSQIP calculator, while six participants (60%) were unfamiliar with the NSQIP calculator.

Grouping

Two out of 16 postoperative complications (13%) were categorized equally between all medical experts and non-experts (Table 5). Surgeons and non-expert grouping were consistent for 11 out of 16 complications (69%); surgeons and clinicians grouping was consistent for 10 out of 16 complications (63%); surgeons and nurses grouping was consistent for 5 out of 16 complications (31%); surgeons and students grouping were consistent for 10 out of 11 complications (63%), and surgeons and administrators were consistent for 6 out of 16 complications (38%). The postoperative complications “Return to OR” and “Death” were complications with the most consistent groupings among experts and non-experts. Postoperative complications with the least consistent grouping were “Pneumonia” and “Colectomy Anastomotic Leak.”

Table 5. Grouping of postoperative complications group

Complications	x=Surgeons (n=1); x=Clinician (n=1); x=Nurses & CNA (n=5); x=Students (n=1); o=Non-experts (n=64)			
	Catastrophic	**Major**	**Moderate**	**Minor**
Serious complication	x x	x x o		
Any complication			x x o	x x
Pneumonia		x o	x x	x
Cardiac complication	x x o	x x		
Surgical site infection		x x o	x x	
Urinary Tract Infection			x	x x x o
Venous Thromboembolism		x x x o	x	
Renal failure	x o	x x x		
Readmission			x x x o	x
Return to OR		x x x x o		
Colectomy Anastomotic Leak	x	x o	x x	
Colectomy Ileus		x o	x x x	
Death	x x x x o			
Discharge to Nursing or rehab Facility			x	x x x o
Sepsis	x x x o	x		

Table 5 Continued

Complications	x=Surgeons (n=1); x=Clinician (n=1); x=Nurses & CNA (n=5); x=Students (n=1); o=Non-experts (n=64)			
	Catastrophic	**Major**	**Moderate**	**Minor**
Extended Length of Hospital Stay			x x o	x x

There were 16 (100%) postoperative complications identified as unfamiliar by non-experts, and 14 (78%) postoperative complications identified as confusing for patients by experts. The experts and non-experts were consistent in the selection of the top three most unfamiliar postoperative complications (Figure 6). Colectomy Anastomotic Leak was the most unfamiliar complication for non-experts (n=51, 80%), followed by Colectomy Ileus (n=48, 75%), Venous Thromboembolism (n=36, 56%) and Renal Failure (n=19, 30%). Venous Thromboembolism was the most frequently selected complication by experts (n=10, 50%), followed by Colectomy Anastomotic Leak (n=3, 40%), Renal Failure (n=4, 40%), and Cardiac Complications (n=4, 40%).

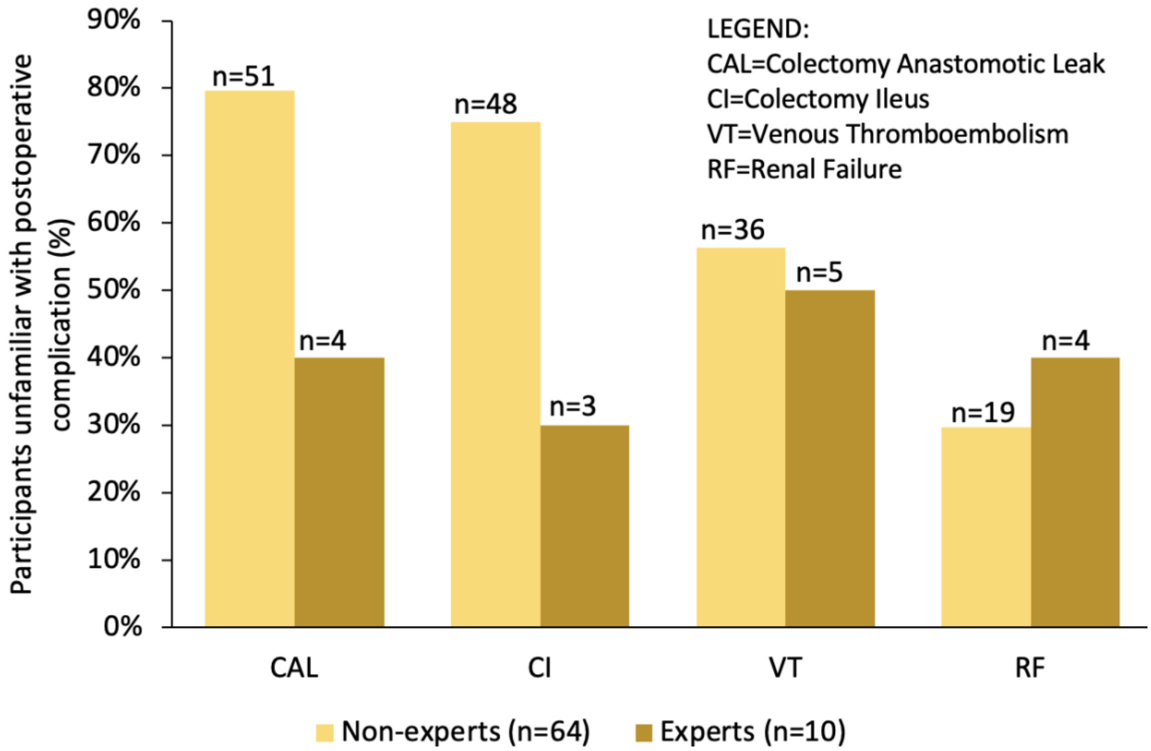


Figure 6. Ranking of unfamiliar complications

Discussion

The study results support the stated hypothesis, that there were significant disparities in the organization and grouping of postoperative complications between different user groups of the NSQIP calculator. Since the goal of this study was to identify the optimal grouping and ranking of postoperative complications, the ranking, and grouping developed by surgeons was considered as "optimal". Unexpectedly, the highest match in grouping results was between surgeons and non-experts (69%). The low match in grouping between surgeons and other healthcare employees (nurses (31%), students (63%), clinicians (63%), etc.) addresses two potential problems of the NSQIP surgical calculator. First, very loosely defined pre-surgical complications allow NSQIP calculator users to self-interpret the exact meaning of complications. Challenges of overgeneralized terms and definitions were addressed by Fowler et. al, who conducted a study to identify how unclear terms impact the outcomes of survey results [80]. According to Fowler, unclearly defined terms are likely to produce biased estimates, and systematic error [80], which was observed in this study.

When expert and non-expert participants were asked to select any postoperative complications, whose understanding could be unclear to the public, experts selected 14 out of 16 complications (87.5%), and non-experts selected 16 out of 16 complications (100%).

To accommodate for a lack of understanding, this study proposed the following layout of postoperative complications in a rank/risk-ordered manner that meets both expert and non-expert users' preferences (Table 6). This new proposed layout of postsurgical complications on the NSQIP user interface provides consistency for the risk of severe complications like death, sepsis, and colectomy anastomotic leak. Where colectomy ileus, urinary tract infection, and discharge to a nursing or rehabilitation facility were located at the bottom of the screen. Furthermore, as a response to the lack of understanding of postoperative complications (Figure 6), and the inconsistency in the ranking (Table 6) additional supporting resources (i.e., an information icon that would allow users to inform themselves about the true meaning of each surgical complication) should be added to the NSQIP calculator user interface next to displayed postoperative complications.

Table 6. Borda Count ranking of postoperative complications.

Complications	**Surgeon**	**Nurses**	**Non-experts**
Death	1	1	1
Sepsis	1	4	4

Table 6 Continued

Complications	Surgeon	Nurses	Non-experts
Colectomy Anastomotic Leak	3	8	7
Serious Complication	4	2	5
Cardiac Complication	5	3	2
Venous Thromboembolism	6	11	6
Return to OR	7	6	8
Renal Failure	8	5	3
Pneumonia	9	9	9
Surgical Site Infection	10	7	10
Readmission	11	12	12
Extended Length of Hospital Stay	12	14	15
Any Complication	13	16	13
Colectomy Ileus	14	10	11
Urinary Tract Infection	15	13	16
Discharge to Nursing or Rehab Facility	16	15	15

Inconsistent postoperative complications grouping results were identified between non-expert users. Zeinab et all. (2016) conducted a study to identify barriers to the implementation of enhanced postoperative recovery [81]. The study identified a lack of patient understanding of health literacy as a challenge that leads to patient noncompliance with recommended care [81]. To overcome this barrier, this study recommends that a new NSQIP calculator UI includes a "clearer" definition of postoperative complications that would allow all types of non-experts to understand displayed content.

A major shortcoming of this case study is the small sample size of expert participants. We believe this study could be improved by surveying a large sample of expert participants. However, since the study was conducted in rural healthcare, where staffing resources are constrained [82], a satisfactory amount of data was analyzed, resulting in valid and transparent conclusions about the current situation.

It is recommended that future researchers expand this study to non-rural healthcare settings with a larger sample population. By doing so, different angles of postoperative understanding could be captured that would provide additional recommendations for NSQIP calculator user interface improvements.

Conclusion

The purpose of this case study was to provide a recommendation for the improvement of the NSQIP calculator user interface, by identifying an optimal order of postoperative complications. Expert and non-expert responses indicated that there is a high level of ambiguity regarding the importance and understanding of postsurgical complications. To overcome the barriers of ambiguity and lack of understanding of health literacy, this case study proposed a new order of postoperative complications on the NSQIP user interface, that was developed by surgeons. Furthermore, this study recommended the addition of a "clear" definition/explanation next to each complication. The major limitation of this study was a small dataset, because of rural healthcare settings. To provide a more robust conclusion, future research should replicate this study in an urban healthcare setting, to capture a larger sample size and provide more robust outcomes.

CHAPTER FOUR

EVALUATION OF DECISION TREE FOR PREDICTING PATIENTS' LENGTH OF STAY AFTER ARTHROPLASTY SURGICAL PROCEDURES IN FRONTIER HEALTHCARE

Contribution of Authors and Co-Authors

Author: Nejc Sitar

Contribution: Conceptualization, Formal Analysis, Methodology, Visualization, Writing

Co-Author: Bernadette McCrory

Co-Author: Faraz Dadgostari

Contribution: Supervision, Conceptualization, Validation, Review and Editing

Co-Author: Bradley Whitaker

Contribution: Conceptualization, Validation, Review and Editing

Manuscript Information

Nejc Sitar, Faraz Dadgostari, Bernadette McCrory, Bradley Whitaker

Proceedings of the Intermountain Engineering, Technology, and Computing Conference (i-ETC)

Status of Manuscript:

☐ Prepared for submission to a peer-reviewed journal
☐ Officially submitted to a peer-reviewed journal
☐ Accepted by a peer-reviewed journal
X Published in a peer-reviewed journal

IEEE Xplore

2023 Intermountain Engineering, Technology, and Computing (IETC)
https://doi.org/10.1109/IETC57902.2023.10152286

Abstract

Predicted inpatient length of stay (LOS) after total joint arthroplasty (TJA) is important information that sets the expectation for the post-surgical recovery period, which helps hospital management plan, and manage surgical resources as efficiently as possible. In this study, the accuracy of prediction models for predicting patients' LOS in a rural community hospital was conducted in the Northwest United States of America. Data were collected from 181 patients for hip, knee, and shoulder surgeries. A decision tree (DT) prediction model was compared with the National Surgical Quality Improvement Program (NSQIP) calculator for predicting the LOS of TJA. The DT model (RMSE = 0.67) provided a more accurate LOS prediction than the NSQIP calculator (RMSE = 1.18). Furthermore, a greater level of interpretability of the decision-making process makes the DT model very applicable in "high stakes" environments like healthcare.

Keywords— Decision tree, NSQIP calculator, length of stay, rural healthcare, perioperative surgical home

Introduction

Total joint arthroplasty (TJA) is one of the most performed elective surgical procedures in the United States of America (USA) [27], [83]. The volume of primary and revised TJAs has dramatically risen over the past decades, to the extent that TJA procedures become relevant to the current political debate regarding cost containment in the American healthcare system [27]. The three most common types of TJA surgeries are hip replacement, knee replacement, and shoulder replacement [28]. The TJA procedures are frequent procedures in the state of Montana, with a higher prevalence percentage for total hip replacement (THR) and total knee replacement (TKR) procedures compared to the national average [29].

The state of Montana is the 4th largest state by square miles in the USA [84], with a population density of 7.4 persons per square mile [84]. As a result of the small population density in the state, there is no operating Level I Trauma care system to support advanced surgical procedures [85]. This study was conducted in a local, rural community hospital and a newly implemented Perioperative Surgical Home (PSH) with a Level III Trauma care system [7]. The PSH is defined by the American Society of Anesthesiologists as "a patient-centered and physician-led multidisciplinary and team-based system of coordinated care that guides the patient throughout the entire surgical experience" [24]. A common tool used by the PSH is the American College of Surgeons – National Surgical Quality Improvement Program (ACS-NSQIP) Surgical Risk predicting calculator [30]. The ACS-NSQIP calculator is a tool that provides patient-specific risk information to guide both surgical decision-making and informed consent [38]. The NSQIP risk calculator was based on the logistic regression model that was developed based on 1.4 million operations performed in over 500 hospitals in the US [38]. One of the predicted outcomes of the ACS-NSQIP risk calculator is also a length of stay (LOS) [38] along with 13 major complications including morbidity and mortality [35]. LOS is one of the most standard performance measures in healthcare, highly related to the likelihood of mortality, the resources used, and morbidity [86].

The increasing availability of electronic health data presents an opportunity for assessing healthcare services [87]. To successfully process and compute complex datasets, Machine learning (ML), the study of tools and methods for identifying data patterns [87] was proposed as an alternative approach for predicting patients' LOS [30]. A decision tree (DT) is a visual and analytical tool in an ML category, primarily used to obtain knowledge in the form of a tree (i.e., flow chart-like structure), where each non-leaf node denotes a test on an attribute, and each branch indicates an output of the test [60]. Due to its hierarchical structure of nodes and branches [60], the DT prediction model was evaluated as an alternative model for predicting LOS in frontier and rural healthcare.

This study builds on the work of Sridhar, Whitaker, Mouat-Hunter, and McCrory (2022), where the authors developed a complex ML random forest (RF) model for predicting TJA patients' LOS in rural community hospitals [30]. To better fit the needs of the clinical setting, a highly interpretable DT model for predicting TJA patients' LOS in rural community hospitals was developed as an alternative to the complex RF model. It was hypothesized that the new DT prediction model would result in a more accurate prediction of LOS (i.e., lower Root Mean Square Error (RMSE) value) and better fit the needs of frontier and rural health systems.

Literature Review

Studies on predicting patients' LOS were recurrent in the current literature [60]. Bacchi, et al. (2020) conducted a review study on the application of ML for the prediction of total hospital inpatient LOS for medical patients [88]. Out of 21 case studies, 10 of them examined LOS in specific medical patient populations, 8 studies were conducted on inpatients in their respective medical centers, and 3 studies included patients with acute kidney injuries, intensive care admission, and elective admission [88]. The study review resulted in variation in ML performance in predicting patients' LOS, primarily due to inconsistent methodological approaches [88].

A study with positive results was conducted on predicting LOS by using ML tools for patients after joint replacement surgeries in a rural community hospital [30]. The objective of the study was to develop an RF model to predict LOS for total joint replacement procedures in rural community hospitals, that would outperform the publicly available NSQIP calculator [30]. Study results have proven the initial hypothesis, that the RF model outperformed the NSQIP calculator for predicting patient LOS in a rural community hospital [30].

After evaluating the current literature on the performance of the ACS-NSQIP Surgical Risk calculator, mixed outcomes have been identified [38], [89]–[92]. Burgess, Smith, Britt, Weireter, and Polk (2017) conducted a study on the ACS-NSQIP calculator predictions for acute care surgery patients, where they identified that the calculator provided surgeons and patients with an objective assessment of the patient’s risk factors including LOS [91]. Contrary, the ACS-NSQIP calculator has proven to be an inadequate model for predicting patients' LOS in patients with major Head and Neck surgeries [92], in patients undergoing Free Flap Reconstruction [90], and for patients undergoing Elective Spine Surgery [89]. Furthermore, Carr, Mears, Barns, and Stambough (2021) performed a study on predicting LOS after the TJA procedure, where they found that actual LOS was significantly shorter than that predicted by the ACS-NSQIP calculator [38].

Methods

Patient Population

The cohort consisted of a rural community hospital with patients undergoing 20 different surgical procedures during the study period of June 6th, 2020, to May 17th, 2022. Out of the 321 included patients, 138 of them were excluded from the dataset as they did not undergo a TJA procedure (44%). The final dataset included 181 patients with hip (n=83, 46%), knee (n=84, 46%), or shoulder (n=14, 8%) procedures.

Dataset

The original dataset consisted of 30 independent variables (e.g., age, gender, comorbidities) and one dependent variable (LOS). 20 independent variables (n=181) were included in the data analysis. Additionally, four independent variables were re-coded into two variables. Specifically, patients' weight and gender were used to calculate a body mass index (BMI), the pre-surgical period was calculated as the duration between the day of admission (DOA) and the day of surgery (DOS). Of the 18 remaining independent variables, 2 were continuous and 16 were categorical.

Missing Values

There were missing values in the original dataset (n=181). Unbalanced data sets can cause negative effects on the performance of ML algorithms like DT [93]. None of the variables had more than 50% of missing values, so no variables were excluded entirely from the dataset [60]. Second, if a continuous variable had less than 12% of missing values, the missing values were replaced by the mean value of that variable [60]. Furthermore, if the variable was categorical, the mode value replaced the missing value [60].

Training and Test Dataset

After cleaning and recoding the data, 181 completed records were extracted and obtained for data mining. Data were separated into training and testing datasets to allow an accurate evaluation of the prediction model [60]. Based on similar studies [30], [94], the data were randomly split at an 80% - 20% ratio to training (n = 144) and testing (n = 37) datasets, respectively. Additionally, the DT model and corresponding prediction models were developed in R computer language, using the Recursive Partitioning and Regression Trees (rpart) package [95].

Model Building

First, patients' LOS was converted from days to hours in the train and test dataset [30]. Next, the regression DT model for predicting patients' LOS was developed on the training dataset (Initial model). To eliminate the nonpredictive parts causing model overfitting, a "pruning" technique was performed on the Initial model [96]. The model pruning was executed by cross-validation procedure, where the optimal model size corresponded to the complexity parameter (cp) with the smallest cross-validation error [97]. The second regression DT model was developed with optimal cp on the training dataset (Final model). The Final model was then used to develop a prediction of LOS on the test dataset. Finally, to allow direct comparison with the NSQIP prediction model, the Final model predictions of LOS and test dataset were converted back from hours to half-day (12-hour) intervals.

NSQIP Calculations

Demographic information of 37 patients in the test dataset was inserted into the NSQIP calculator, which calculated the predicted LOS in half-day (12-hour) intervals.

Model Validation

The statistical validity of the prediction models was tested by using the two-sided t-test at ($\alpha \leq 0.05$). Next, the prediction models' goodness-of-fit was tested by the Root-Mean-Square Error (RMSE), the Mean Absolute Error (MAE) [98], and the Mean Absolute Percentage Error (MAPE) [99]. The RMSE, MAE, and MAPE behave like a cost function (i.e., the lower the value, the better the model is performing) [30].

Results

Statistical Analysis

First, descriptive statistical analyses on the train and test datasets were performed. Train and test datasets were similarly represented in all variables. Most patients were females (63%), and most patients belonged to the age group between 65-74 (43%). A large proportion of patients had Mild Systematic Diseases [100] (67%), and a high percentage of patients listed Medicare as their primary insurance (68%). There was very little difference in the Duke Activity Status Index (DASI) [101] between the train and test datasets. The average LOS was larger in train datasets compared to the test dataset, but because the difference was less than a half day (7 hours) no extreme values were excluded from the training dataset. Additionally, there were no patients with partially dependent functional status, insulin prescriptions, or dyspnea diagnosis identified in the test dataset. However, since patients with these conditions represented a very small proportion of the entire dataset (n=181), 2%, 1.5%, and 1.5% respectively, the test dataset was still an adequate tool to evaluate prediction model performance.

Decision Tree

The Initial regression DT model developed from the training dataset resulted in the DT diagram with 6 chance nodes and 8 end nodes. The DASI was the most significant variable in the DT model. Additionally, Zip Code and Income had a variable importance index of around 10 and contributed to the DT model as well. To identify the optimal DT model, a model pruning technique was implemented [102]. The optimal complexity parameter (cp) value was identified from cross-validation error versus the cp value and tree size plot. The cross-validation error was minimized at the cp value 0.0416 and the size of the tree was equal to two, as could be seen in Figure 7.

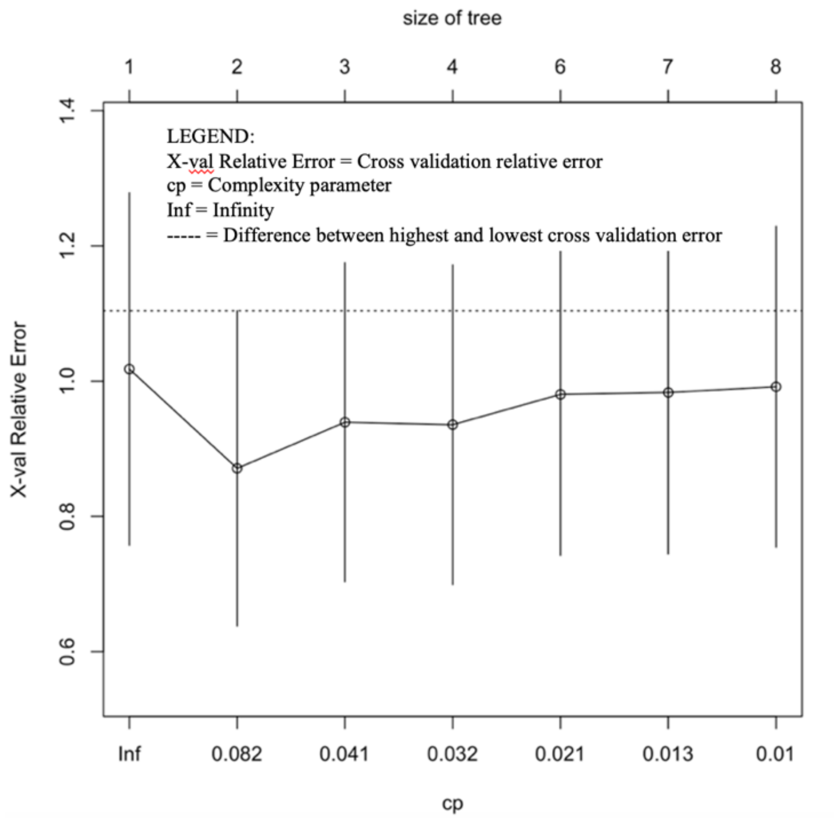


Figure 7. The cross-validation error versus cp value and tree size plot

A Final DT model that can be seen in Figure 8 was built on the training dataset, including an optimal cp value (cp =0.0416). According to the model, 55% of patients with an average LOS equal to 32 hours had a DASI index greater or equal to 5.1, and an annual income greater or equal to $49,000. 38% of patients with an average LOS equal to 45 hours had a DASI index greater or equal to 5.1 and an annual income of less than $49,000. Finally, 7% of patients with an average LOS equal to 82 hours had a DASI index of less than 5.1.

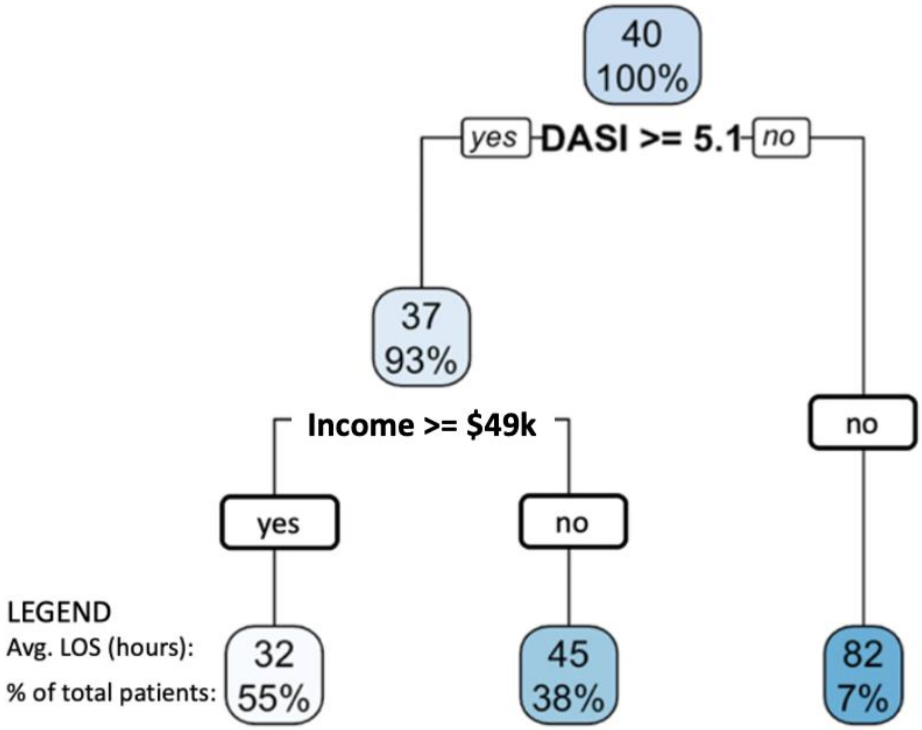


Figure 8. Final DT model diagram: The color intensity corresponds to the predicted LOS (Darker the color, the longer the LOS [103])

The DASI index was the most significant variable in the Final DT model (Figure 9). Zip Code, Income, and Functional Status had a variable importance index of approximately 10 and contributed to the DT model as well. Furthermore, independent variables not listed in Figure 9, resulted in a variable importance index of less than 1 (i.e., no significant contribution to the Final model).

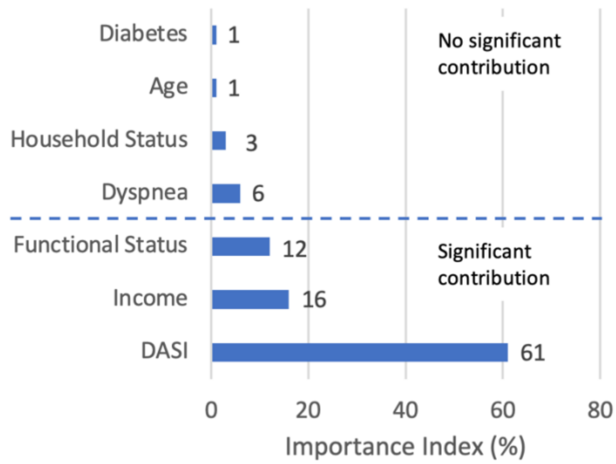


Figure 9. Final DT Model variable importance index (%) vs. independent variables

Model Validation

Predictions developed by the Initial DT model, Final DT model, and NSQIP calculator were compared with the patient's actual LOS and evaluated with RMSE, MAE, MAPE, and two-sided t-tests. Predictions developed by all tree models satisfied the 95% level of significance (Table 7). All three models resulted in a MAPE value of approximately 5%. Finally, the Final DT model outperformed all other models in both RMSE and MAE performance measures, while the NSQIP model produced the least desirable results for both RMSE and MAE.

Table 7 Prediction models validation results

Model validation	**RMSE**	**MAE**	**MAPE**	**t statistic (dof = 36), p-value**
Initial Model Prediction	18.24	14.64	5%	48.528, 0.05
Final Model Prediction	16.08	12.96	5%	99.024, <0.00
NSQIP Model Prediction	28.32	24	5%	195.36, <0.00

Initial and Final DT models' prediction outputs resulted in a mean LOS value of 39.12 hours and 42.48 hours respectively, which was relatively close to the Actual LOS with the corresponding mean value of 33.36 hours (Figure 10). The NSQIP prediction of LOS deviated from the other three values, with the LOS equal to 56.16 hours.

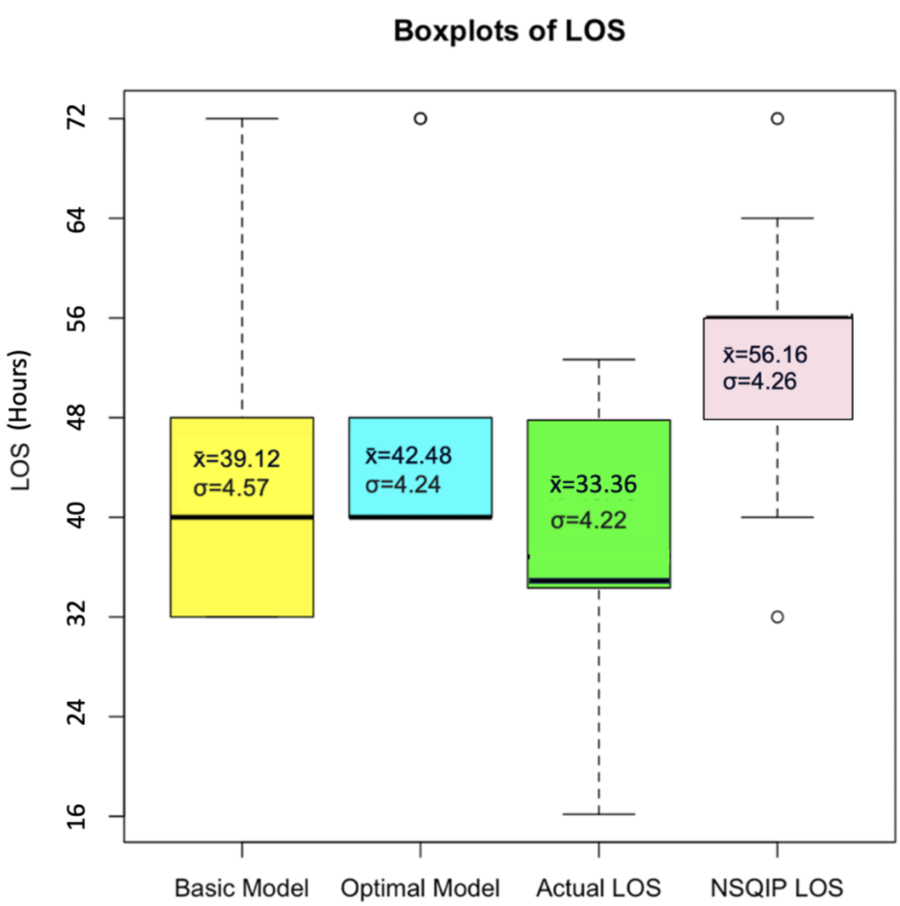


Figure 10. Predicted LOS

Discussion

Preliminary Findings

The decision Tree model resulted in a more accurate prediction of patients' LOS than the NSQIP calculator, as it outperformed the NSQIP calculator in two out of three goodness-of-fit performance indicators (Table 7). Based on the RMSE, the optimal ML performance measure for predicting LOS [30], the DT model (RMSE = 16.08) outperformed the NSQIP calculator (RMSE = 28.32).

The difference in the quality of prediction between DT and NSQIP models can be explained by the following factors. First, the DT model was based on the dataset (n=144) of patients in frontier areas, while the NSQIP calculator was based on the data from over five million surgeries performed at 885 hospitals around the United States of America [38]. Despite the significantly smaller sample size, the superiority of the DT model indicates the need for rural healthcare demographic-specific prediction models.

The patients included in this study were predominantly females representing 63% of the patients, which was 3% higher compared to the NSQIP database [104]. The proportion of hip and knee surgeries in frontier healthcare was almost the same, which was not the case at the national level, where hip surgeries accounted for an additional 8% of total knee surgeries [105]. Furthermore, the proportion of Medicare patients having a TJA in rural healthcare is 4.5% higher compared to the national average [106].

Next, complex ML tools have already been proven as a superior prediction method for predicting patients' LOS compared to the NSQIP calculator [30]. However, the results of this study indicate that even the most basic ML tools, like DT, provide better prediction of LOS for TJA patients than the logistic regression algorithm used by the NSQIP calculator [38]. Positive prediction outcomes of ML tools indicate future opportunities for healthcare improvements [88]. However, ML models require adequate data to develop prediction models [107]. Collecting and implementing data to facilitate the quality of care provided in rural healthcare is an opportunity that this study attempted to address and is the focus of future research.

Finally, the greatest advantage of the DT model compared to the NSQIP calculator is its high level of interpretability [108]. The ability to understand prediction output and its corresponding decision-making process is critical for practical use, especially in high-risk environments like healthcare [108]. The tree-like structure of the DT diagram clearly highlights the branching variables with corresponding values, which allows a clear understanding and visualization of prediction outcomes and the decision-making process [38]. The Final DT model consisted of only three branches, where the most important independent variable (Importance index = 61%) was the Duke Activity Status Index (DASI) (Figure 8). In clinical practice, DASI is used to assess medical treatments and assist clinical decisions, while in control clinical trials, the DASI can serve to evaluate interventions and support a cost/benefit analysis of the treatment [101]. According to this model, the high importance of DASI for predicting LOS is very valuable information in the preoperative stage, as it allows clinicians to prioritize patients' activity status to reduce the likelihood of extended LOS. On another hand, the NSQIP calculator output included the predicted value of LOS and the list of independent variables that have contributed to the model prediction [109]. However, the NSQIP calculator did not provide any insight into the decision-making process [35], which made the model difficult to understand and interpret.

To summarize, the NSQIP calculator provided a less accurate prediction of LOS with a lower level of interpretability, which indicates that a DT model was a better prediction tool for predicting patients' LOS in the rural healthcare environment.

Study Limitations and Future Opportunities

The biggest limitation of this study was the robustness of the developed models. The challenge of model robustness is especially known for DT models, as a very slight change in the dataset can lead to a different tree diagram [110]. Since the Final DT model was based on 144 data points and evaluated with 37 data points, it is very sensitive to the changes in independent variables. Furthermore, the Final DT model produced only three predicted LOS, which constrained the model predictions. As a result of that, our model should be used as an informative tool, allowing clinicians to develop a better understanding of factors that contributed to an extended LOS. However, despite these limitations, the model developed in this study is completely relevant and statistically evaluated. Furthermore, since rural healthcare deals with a significantly lower percentage of patients compared to the major clinical centers [111], this study represents a good assessment of rural healthcare and could hopefully lead to beneficial improvements for patients, and clinicians.

Future studies should focus on replicating this study with a larger dataset to yield a more robust and less sensitive model. Additionally, future researchers should evaluate the interpretability of the DT model with physicians and their level of agreement with the DT-provided predicted LOS.

Conclusion

The DT model provided a more accurate prediction of LOS compared to the NSQIP calculator for patients undergoing TJA procedures in rural community hospitals. The DT model output displayed the decision-making process and emphasized the importance of DASI for model prediction. The difference in the demographic characteristics of study participants compared to the national NSQIP database emphasized the need for a rural healthcare-specific prediction model and consequently the need for data collection in rural healthcare. In alignment with existing research work, this study proved the superiority of ML models for predicting TJA patients' LOS in rural healthcare.

CHAPTER FIVE

CONCLUSION, RECOMMENDATIONS FOR FUTURE WORK, AND OTHER

Conclusion

This thesis addressed the barriers of the PSH clinic in the Northwest. In particular, the main focus was on the "Decision for Surgery," "Preoperative," and "Postoperative" stages of the PSH model (Figure 11). The NSQIP calculator UI was analyzed, as it is the currently used prediction tool of the PSH, to improve the interpretability and shared understanding of displayed output. This research revealed disparities in the organization of postoperative complications not just between expert and non-expert participants (69% match), but also between different groups of expert participants (i.e., a match between surgeons and nurses was 31%). Open-ended definitions (i.e., lacking concrete definitions presented in the UI) of postoperative complications led study participants to self-interpret the meaning and severity of the UI's presented complications [80]. To overcome this barrier, a research team proposed a novel layout of postoperative complications in a combined novice- and expert-prescribed descending order of complication severity (Table 6). This recommended novel UI layout graphically prioritizes postoperative complications by severity to reduce the likelihood of misinterpretation among the PSH clinical team and patient for optimized decision-making. This novel UI layout facilitates longitudinal patient care, particularly in the preoperative stage of the PSH model, but can also be used in other phases of the PSH model for continued risk monitoring/assessment (Figure 11). To quickly access definitions of each postoperative complication additional NSQIP UI functionalities were recommended. These suggestions will increase the ease of understanding of the NSQIP calculator output, allowing for better decision-making in the “Decision for Surgery” stage of the PSH model (Figure 11).

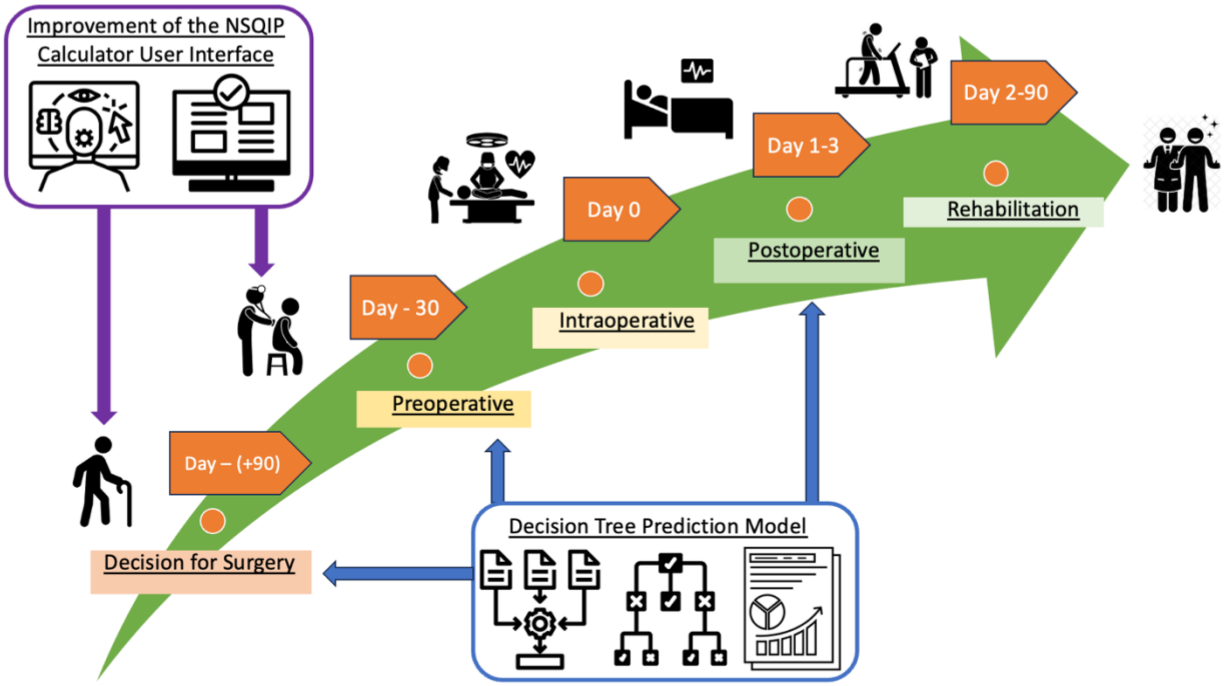


Figure 11. Proposed improvements of the PSH model (add a link to the human)

From the result, the DT model resulted in a more accurate prediction of TJA patients’ LOS compared to the NSQIP calculator (Table 7). The DT model resulted in stronger prediction performance compared to the NSQIP calculator for the following reasons. First, the DT model was based on the dataset of patients living in frontier/rural areas (n=144), while the NSQIP calculator was based on the data collected at 885 hospitals around the United States of America [38]. Due to the differences in demographic characteristics of a rural population compared to the national database, the DT model provided a more accurate prediction of LOS for patients in rural areas despite the significantly smaller dataset. Furthermore, the NSQIP database includes insufficient numbers of orthopedic procedures (i.e., only 12% of orthopedic patient data contributed to the analysis), while the DT model is entirely based on the arthroplasty patient data [112]. As a result, the NSQIP calculator predicted LOS higher than the actual LOS for TJA performed in the rural PSH clinic, which was consistent with existing research [30]. This outcome indicates the importance of developing rural-specific prediction models in healthcare that would better capture the characteristics of the rural population.

Next, complex ML models outperformed the NSQIP calculator in predicting LOS [30]. This study went further than existing research and demonstrated that even the most basic ML model such as DT outperformed the NSQIP calculator despite being based on a small data set (n=144). The main advantage of the DT model is not only in the prediction accuracy but also in its high level of model interpretability [108]. The tree-like structure of the DT diagram exposed branching variables with corresponding values, which allows for a better understanding of the decision-making process [38]. As seen in Figure 9, The Duke Activity Status Index (DASI) is a very significant variable that contributes to prolonged LOS. By prioritizing patients' DASI, clinicians can significantly impact the likelihood of extended LOS for TJA patients in rural healthcare. The results from this study recommend the inclusion of the DASI in the NSQIP calculator as one of the prediction variables. Furthermore, due to DASI's high importance in rural healthcare, it is recommended that DASI be included in the preoperative stage of arthroplasty procedures, as one of the key indicators to assess patients' readiness for the surgical procedure.

The improved prediction accuracy and interpretability of the DT model compared to the NSQIP calculator will positively impact the decision-making in the “Decision for Surgery,” “Preoperative,” and “Postoperative” stages of the PSH model. More accurate prediction of LOS will allow patients living in rural areas and a PSH team to jointly make a better decision on the surgery recommendation (Figure 11) [30]. In the preoperative stage of the PSH model, higher interpretability of the DT prediction better facilitates the PSH team to identify patients' conditions that impact hospital LOS. With that knowledge, the PSH team will prepare a plan for treating patients' conditions during a preoperative process, with the goal of reducing the likelihood of extended hospital LOS and any other surgical and post-surgical complications (**Error! Reference source not found.**). Finally, a more accurate prediction of LOS would allow hospitals to better plan their resources and optimize their operations in the postoperative stage of the PSH model, such as patient scheduling, daily staffing, and bed allocation (Figure 11).

Recommendation for Future Work

To improve users' experience with the NSQIP calculator, a pop-up definition callout hyperlinked to each postoperative complication is proposed. Future researchers should redesign the NSQIP calculator UI, by enabling users to access a definition of postoperative complication by "clicking" on a postoperative complication name (Figure 12). An expected outcome of added functionality is a reduction in the number of participants who did not understand the meaning of postoperative complications.

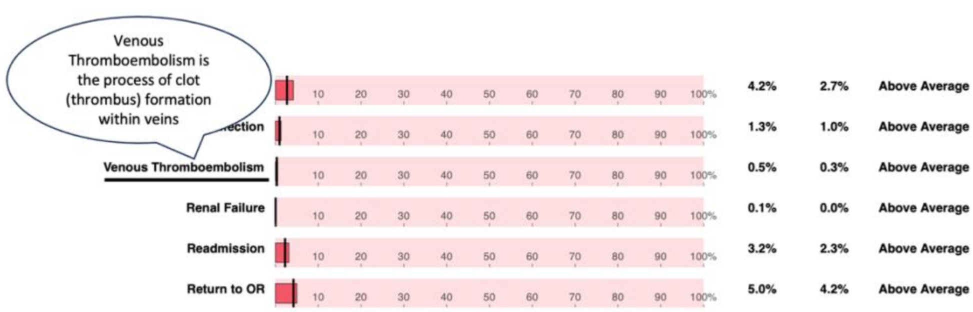


Figure 12. Proposed functionality improvement of NSQIP calculator UI [113]

A major shortcoming of this thesis is the small sample size of study participants. According to Jacob Nielsen, testing the usability of a UI with 5 participants results in similar outcomes as the test would be completed having many more participants [72]. However, when performing a card sorting study, Nielsen recommended 15 participants per user group [72].

This usability case study consisted of 10 expert participants, which is below the Nielsen recommendation. Future research should focus on replicating the study with a larger pool of expert participants (n>15) to overcome study limitations. Furthermore, replicating a case study in an urban setting would allow researchers to analyze differences in the understanding of postoperative complications between urban and rural populations. Despite a smaller sample size, this study gives a preliminary representation of the participants understanding of postoperative complications in rural healthcare. Based on our knowledge, no other research has been performed that analyzes expert and non-expert understanding of postoperative complications in rural healthcare.

Future research should evaluate the interpretability and prediction accuracy of the DT and RF models in clinical settings. Complex ML models like RF are known to provide high accuracy of prediction, but not end-to-end interpretability of its outputs [114]. On the other hand, simple ML models such as DT provide a better interpretation of outputs but produce less accurate predictions [114]. When estimating the interpretability of the two models with clinicians and combined with the model's estimated accuracy, a prediction algorithm per specific clinical environment should be constructed. A limitation of the DT model for predicting LOS of the TJA patients was the robustness of the model due to its small sample size (n=144). Such limitation is a common shortcoming of many DT models, where even a minor deviation in the dataset can cause a significant shift in the model’s outcome [110].

Another opportunity to progress this work is to focus on replicating this study using a larger dataset to deepen our understanding and develop a more robust model. It is estimated that the sample size should increase to about 16,000 participants [115]. One study identified through a progressive sampling methodology that the 16,000 data points universally resulted in an optimal DT performance [115]. Despite the small sample size used in this study, the current DT model provided a stronger prediction accuracy of the TJA LOS compared to the NSQIP calculator (Table 7). The presented DT model indicates that PSH is making a positive impact on a rural healthcare environment and demonstrates the need for expanded data collection efforts in rural healthcare.

Inclusivity in Healthcare

I would like to dedicate the final paragraph of this thesis to addressing the underrepresentation of minority groups in clinical leadership positions in the US. Persons underrepresented in medicine (UIM) include African Americans, Hispanics/Latinos, Native Americans, Alaskans, Hawaiians, and mainland Puerto Ricans [116]. In the US Northwest, approximately 6.5% of the population identifies themselves as American Indians and Alaska Natives [25]. However, as reported by the American Medical Association, American Indians and Alaska Natives represent only 0.4% of the physician workforce [117]. As a response to the lack of diversity among physicians, Sola and the research team proposed the implementation of mentorship to mitigate this problem [116]. Mentorship is defined as a multidimensional approach of intentional recruitment, accessible equitable resources, and opportunities for mentoring support that must start at the beginning of the pipeline with pre-medical students [116]. Mentoring was recognized as a valuable methodology for increasing diversity, equity, and inclusion within the surgical team, thus improving the status of UIMs. Furthermore, increasing representation of UIMs positively impacted patients' surgical care, which addressed the importance of diversification and the need for a larger representation of underrepresented groups occupying physician positions [116]. By addressing the problem of diversification in every industry, not just healthcare, I believe we can improve the well-being of many citizens, and provide equal opportunities for every American and non-American like myself, to achieve their full potential.

REFERENCES CITED

[1] U.S. Census Bureau, “Rural America,” U. S. Department of Commerce. Accessed: Jul. 14, 2023. [Online]. Available: https://libanswers.csudh.edu/faq/377312

[2] Economic Research Service, “Rural-Urban Commuting Area Codes,” U.S. DEPARTMENT OF AGRICULTURE. Accessed: Jul. 14, 2023. [Online]. Available: https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes/documentation/

[3] I. S. Moscovice, S. Moscovice, and A. Director, “Rural Hospitals: A Literature Synthesis and Health Services Research Agenda Agricultural counties; manufacturing counties; mining, oil, and energy counties; and persistent poverty counties account for more than.”

[4] American Hospital Association, “U.S. Rural Hospitals,” American Hospital Association.

[5] G. Mark Holmes, George H. Pink, Sarah A. Friedman, and Hilda A. Howard, “A Comparison of Rural Hospitals with Special Medicare Payment Provisions to Urban and Rural Hospitals Paid Under Prospective Payment,” Aug. 2010. Accessed: Jul. 30, 2023. [Online]. Available: https://www.shepscenter.unc.edu/wp-content/uploads/2014/10/FR98.pdf

[6] American Trauma Society, “Trauma Center Levels Explained.” Accessed: Jul. 14, 2023. [Online]. Available: https://www.amtrauma.org/page/traumalevels

[7] Alyssa Johnson and Carol Kussman, “Trauma Facility Designation Levels,” MONTANA.GOV OFFICIAL STATE WEBSITE. Accessed: Jul. 14, 2023. [Online]. Available: https://dphhs.mt.gov/assets/publichealth/EMSTS/Trauma/MTDesignatedFacilityListMarch2021.pdf

[8] T. Kotzee and ID. Couper, “What interventions do South Africanqualified doctors think will retain them in rural hospitals of the Limpopo province ofSouth Africa?,” *The International Electronic Journal of Rural and Remote Healthcare*, Sep. 2006.

[9] R. Carter Clement, S. B. Bhat, M. E. Clement, and J. C. Krieg, “Medicare reimbursement and orthopedic surgery: past, present, and future,” *Current Reviews in Musculoskeletal Medicine*, vol. 10, no. 2. Humana Press Inc., pp. 224–232, Jun. 01, 2017. doi: 10.1007/s12178-017-9406-7.

[10] Center for Medicare & Medicaid Services, “Value-Based Purchasing Programs,” Home Health Agency. Accessed: Jul. 15, 2023. [Online]. Available: https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/value-based-programs/value-based-programs

[11] H.-C. Liu, “FMEA for Proactive Healthcare Risk Analysis: A Systematic Literature Review,” in *Improved FMEA Methods for Proactive Healthcare Risk Analysis*, Springer Singapore, 2019, pp. 15–45. doi: 10.1007/978-981-13-6366-5_2.

[12] T. Brackett, L. Comer, and R. Whichello, “Do Lean Practices Lead to More Time at the Bedside?,” *Journal for Healthcare Quality*, vol. 35, no. 2, pp. 7–14, Mar. 2013, doi: 10.1111/j.1945-1474.2011.00169.x.

[13] E. P. Tagge, A. S. Thirumoorthi, J. Lenart, C. Garberoglio, and K. W. Mitchell, “Improving operating room efficiency in academic children’s hospital using Lean Six Sigma methodology,” *J Pediatr Surg*, vol. 52, no. 6, pp. 1040–1044, Jun. 2017, doi: 10.1016/j.jpedsurg.2017.03.035.

[14] C. R. A. Hallam and C. Contreras, “Lean healthcare: scale, scope and sustainability,” *Int J Health Care Qual Assur*, vol. 31, no. 7, pp. 684–696, Aug. 2018, doi: 10.1108/IJHCQA-02-2017-0023.

[15] T. Ohno and N. Bodek, *Toyota Production System*. Productivity Press, 2019. doi: 10.4324/9780429273018.

[16] C. R. A. Hallam and C. Contreras, “Lean healthcare: scale, scope and sustainability,” *Int J Health Care Qual Assur*, vol. 31, no. 7, pp. 684–696, Aug. 2018, doi: 10.1108/IJHCQA-02-2017-0023.

[17] B. Klefsjö, H. Wiklund, and R. L. Edgeman, “Six sigma seen as a methodology for total quality management,” *Measuring Business Excellence*, vol. 5, no. 1, pp. 31–35, Mar. 2001, doi: 10.1108/13683040110385809.

[18] Forrest W. Breyfogle III, *Implementing Six Sigma, Smarter Solutions Using Statistical Methods*, 2nd ed. John Wiley & Sons, Inc., 2003.

[19] J. Antony, P. Palsuk, S. Gupta, D. Mishra, and P. Barach, “Six Sigma in healthcare: a systematic review of the literature,” *International Journal of Quality & Reliability Management*, vol. 35, no. 5, pp. 1075–1092, May 2018, doi: 10.1108/IJQRM-02-2017-0027.

[20] S. E. Mason, C. R. Nicolay, and A. Darzi, “The use of Lean and Six Sigma methodologies in surgery: A systematic review,” *The Surgeon*, vol. 13, no. 2, pp. 91–100, Apr. 2015, doi: 10.1016/j.surge.2014.08.002.

[21] O. Desebbe, T. Lanz, Z. Kain, and M. Cannesson, “The perioperative surgical home: An innovative, patient-centred and cost-effective perioperative care model,” *Anaesth Crit Care Pain Med*, vol. 35, no. 1, pp. 59–66, Feb. 2016, doi: 10.1016/j.accpm.2015.08.001.

[22] B. A. Kash, Y. Zhang, K. M. Cline, T. Menser, and T. R. Miller, "The perioperative surgical home (PSH): A comprehensive review of us and non-us studies shows predominantly positive quality and cost outcomes," *Milbank Quarterly*, vol. 92, no. 4. pp. 796–821, Dec. 01, 2014. doi: 10.1111/1468-0009.12093.

[23] "The IHI Triple Aim," 2023. Accessed: Jan. 21, 2023. [Online]. Available: https://www.ihi.org/Engage/Initiatives/TripleAim/pages/default.aspx

[24] Z. N. Kain *et al.*, "The Perioperative Surgical Home as a Future Perioperative Practice Model," *Anesth Analg*, vol. 118, no. 5, pp. 1126–1130, May 2014, doi: 10.1213/ANE.0000000000000190.

[25] U.S. Census Bureau, "Quick Facts Montana," U.S. Department of Commerce. Accessed: Jul. 18, 2023. [Online]. Available: https://www.census.gov/quickfacts/fact/table/MT/PST045222

[26] A. A. Norful, K. de Jacq, R. Carlino, and L. Poghosyan, "Nurse practitioner–physician comanagement: A theoretical model to alleviate primary care strain," *Ann Fam Med*, vol. 16, no. 3, pp. 250–256, May 2018, doi: 10.1370/afm.2230.

[27] M. Sloan, A. Premkumar, and N. P. Sheth, "Projected Volume of Primary Total Joint Arthroplasty in the U.S., 2014 to 2030," *Journal of Bone and Joint Surgery*, vol. 100, no. 17, pp. 1455–1460, Sep. 2018, doi: 10.2106/JBJS.17.01617.

[28] "Types of Arthroplasty," Stanford Medicine.

[29] H. M. Kremers *et al.*, "Prevalence of total hip and knee replacement in the United States," *Journal of Bone and Joint Surgery - American Volume*, vol. 97, no. 17, pp. 1386–1397, Sep. 2014, doi: 10.2106/JBJS.N.01141.

[30] S. Sridhar, B. Whitaker, A. Mouat-Hunter, and B. McCrory, "Predicting Length of Stay using machine learning for total joint replacements performed at a rural community hospital," *PLoS One*, vol. 17, no. 11, p. e0277479, Nov. 2022, doi: 10.1371/journal.pone.0277479.

[31] S. F. Khuri *et al.*, "The Department of Veterans Affairs' NSQIP," *Ann Surg*, vol. 228, no. 4, pp. 491–507, Oct. 1998, doi: 10.1097/00000658-199810000-00006.

[32] P. Parikh *et al.*, "Pancreatectomy risk calculator: an ACS-NSQIP resource," *HPB*, vol. 12, no. 7, pp. 488–497, Sep. 2010, doi: 10.1111/j.1477-2574.2010.00216.x.

[33] S. Klemm, S. van Broeckhuysen-Kloth, S. van Vliet, L. Oosterhuis, and R. Geenen, "Personalized treatment outcomes in patients with somatoform disorder: A concept mapping study," *J Psychosom Res*, vol. 109, pp. 19–24, Jun. 2018, doi: 10.1016/j.jpsychores.2018.03.009.

[34] K. Y. Bilimoria *et al.*, "Development and Evaluation of the Universal ACS NSQIP Surgical Risk Calculator: A Decision Aid and Informed Consent Tool for Patients and Surgeons," *J Am Coll Surg*, vol. 217, no. 5, pp. 833-842e3, Nov. 2013, doi: 10.1016/j.jamcollsurg.2013.07.385.

[35] "ACS NSQIP Calculator," American College of Surgeons National Surgical Quality Improvement Program. Accessed: Jun. 15, 2023. [Online]. Available: https://riskcalculator.facs.org/RiskCalculator/PatientInfo.jsp

[36] G. Scotton *et al.*, "Is the ACS-NSQIP Risk Calculator Accurate in Predicting Adverse Postoperative Outcomes in the Emergency Setting? An Italian Single-center Preliminary Study," *World J Surg*, vol. 44, no. 11, pp. 3710–3719, Nov. 2020, doi: 10.1007/s00268-020-05705-w.

[37] M. E. Cohen, Y. Liu, J. B. Liu, C. Y. Ko, and B. L. Hall, "Use of a Single CPT Code for Risk Adjustment in American College of Surgeons NSQIP Database: Is There Potential Bias with Practice-Pattern Differences in Multiple Procedures under the Same Anesthetic?," *J Am Coll Surg*, vol. 226, no. 3, pp. 309–316, Mar. 2018, doi: 10.1016/j.jamcollsurg.2017.12.020.

[38] C. J. Carr, S. C. Mears, C. L. Barnes, and J. B. Stambough, "Length of Stay After Joint Arthroplasty is Less Than Predicted Using Two Risk Calculators," *J Arthroplasty*, vol. 36, no. 9, pp. 3073–3077, Sep. 2021, doi: 10.1016/j.arth.2021.04.010.

[39] P. Dotson, "CPT ® Codes: What Are They, Why Are They Necessary, and How Are They Developed?," *Adv Wound Care (New Rochelle)*, vol. 2, no. 10, pp. 583–587, Dec. 2013, doi: 10.1089/wound.2013.0483.

[40] A. Must and S. E. Anderson, "Body mass index in children and adolescents: Considerations for population-based applications," *International Journal of Obesity*, vol. 30, no. 4. pp. 590–594, Apr. 2006. doi: 10.1038/sj.ijo.0803300.

[41] C. L. Nelson, J. E. Herndon, D. B. Mark, D. B. Pryor, R. M. Califf, and M. A. Hlatky, "Relation of Clinical and Angiographic Factors to Functional Capacity as Measured by the Duke Activity Status Index," 1979.

[42] Mary Louise, Lane Carlson, and John Kumar, "Engaging Patients in Managing Their Health Care: Patient Perceptions of the Effect of a Total Joint Replacement Presurgical Class," *Perm J*, vol. 16, no. 3, pp. 42–47, 2012, Accessed: Feb. 02, 2023. [Online]. Available: https://www.thepermanentejournal.org/doi/pdf/10.7812/tpj.16.issue-3#page=44

[43] H. Krampe, A. Barth-Zoubairi, T. Schnell, A. L. Salz, L. F. Kerper, and C. D. Spies, "Social Relationship Factors, Preoperative Depression, and Hospital Length

of Stay in Surgical Patients," *Int J Behav Med*, vol. 25, no. 6, pp. 658–668, Dec. 2018, doi: 10.1007/s12529-018-9738-8.

[44] T. H. Grubesic, "Zip codes and spatial analysis: Problems and prospects," *Socioecon Plann Sci*, vol. 42, no. 2, pp. 129–149, Jun. 2008, doi: 10.1016/j.seps.2006.09.001.

[45] C. Cuadrado, F. Crispi, M. Libuy, G. Marchildon, and C. Cid, "National Health Insurance: A conceptual framework from conflicting typologies," *Health Policy (New York)*, vol. 123, no. 7, pp. 621–629, Jul. 2019, doi: 10.1016/j.healthpol.2019.05.013.

[46] T. Murakami *et al.*, "Characterization of predictors of in-hospital cardiac complications of takotsubo cardiomyopathy: Multi-center registry from Tokyo CCU Network," *J Cardiol*, vol. 63, no. 4, pp. 269–273, Apr. 2014, doi: 10.1016/j.jjcc.2013.09.003.

[47] J. Rumball-Smith and P. Hider, "THE NEW ZEALAND MEDICAL JOURNAL The validity of readmission rate as a marker of the quality of hospital care, and a recommendation for its definition," 2009.

[48] "The Definition of Death," *Stanford Encyclopedia of Philosophy*, Oct. 2007, Accessed: Feb. 07, 2023. [Online]. Available: http://seop.illc.uva.nl/entries/death-definition/

[49] I. Matot and C. L. Sprung, "Definition of sepsis," *Intensive Care Med*, vol. 27, no. 14, pp. S3–S9, Feb. 2001, doi: 10.1007/PL00003795.

[50] K. Stone, R. Zwiggelaar, P. Jones, and N. mac Parthaláin, "A systematic review of the prediction of hospital length of stay: Towards a unified framework," *PLOS Digital Health*, vol. 1, no. 4, p. e0000017, Apr. 2022, doi: 10.1371/journal.pdig.0000017.

[51] Philippe Palanque and Fabio Paternò, *Interactive Systems*, vol. 1946. Limerick: 7th InternationalWorkshop, 2000.

[52] Nuno Jardim Nunes and Joao Falcao e Cunha, "Wisdom – A UML Based Architecture for Interactive Systems," *Lecture Notes in Computer Science*, pp. 191–105, 2001.

[53] J. Nielsen, *Usability Engineering*. Boston: AP Professional, 1993.

[54] S. Hannah, "Sorting Out Card Sorting: Comparing Methods for Information Architects, Usability Specialists, and Other Practitioners," 2005.

[55] J. R. Wood and L. E. Wood, “Card Sorting: Current Practices and Beyond History and Assumptions,” 2008.

[56] S. Fincher and J. Tenenberg, “Making sense of card sorting data,” *Expert Syst*, vol. 22, no. 3, pp. 89–93, Jul. 2005, doi: 10.1111/j.1468-0394.2005.00299.x.

[57] P. Emerson, “The original Borda count and partial voting,” *Soc Choice Welfare*, vol. 40, no. 2, pp. 353–358, Feb. 2013, doi: 10.1007/s00355-011-0603-9.

[58] S. Dash, S. K. Shakyawar, M. Sharma, and S. Kaushik, “Big data in healthcare: management, analysis and future prospects,” *J Big Data*, vol. 6, no. 1, p. 54, Dec. 2019, doi: 10.1186/s40537-019-0217-0.

[59] Z. Obermeyer and E. J. Emanuel, “Predicting the Future — Big Data, Machine Learning, and Clinical Medicine,” *New England Journal of Medicine*, vol. 375, no. 13, pp. 1216–1219, Sep. 2016, doi: 10.1056/NEJMp1606181.

[60] P. R. Hachesu, M. Ahmadi, S. Alizadeh, and F. Sadoughi, “Use of Data Mining Techniques to Determine and Predict Length of Stay of Cardiac Patients,” *Healthc Inform Res*, vol. 19, no. 2, p. 121, 2013, doi: 10.4258/hir.2013.19.2.121.

[61] H. Maradit Kremers *et al.*, “Prevalence of Total Hip and Knee Replacement in the United States,” *The Journal of Bone and Joint Surgery-American Volume*, vol. 97, no. 17, pp. 1386–1397, Sep. 2015, doi: 10.2106/JBJS.N.01141.

[62] S. Mohanty *et al.*, “Risk of Discharge to Postacute Care,” *JAMA Surg*, vol. 150, no. 5, p. 480, May 2015, doi: 10.1001/jamasurg.2014.3176.

[63] K. Y. Bilimoria *et al.*, “Development and Evaluation of the Universal ACS NSQIP Surgical Risk Calculator: A Decision Aid and Informed Consent Tool for Patients and Surgeons,” *J Am Coll Surg*, vol. 217, no. 5, pp. 833-842e3, Nov. 2013, doi: 10.1016/j.jamcollsurg.2013.07.385.

[64] M. Green, “The University of Alberta user interface management system,” *ACM SIGGRAPH Computer Graphics*, vol. 19, no. 3, pp. 205–213, Jul. 1985, doi: 10.1145/325165.325286.

[65] P. E. Sigmund, *Nicholas of Cusa and Medieval Political Thought*. Harvard University Press, 1963. doi: 10.4159/harvard.9780674433427.

[66] G. Mackenzie, “The definition and classification of pneumonia,” *Pneumonia*, vol. 8, no. 1, Dec. 2016, doi: 10.1186/s41479-016-0012-z.

[67] N. Alhammadi *et al.*, “Public awareness of diabetes complications and its effect on treatment compliance in Asir region, Saudi Arabia,” *J Family Med Prim Care*, vol. 11, no. 11, p. 6812, 2022, doi: 10.4103/jfmpc.jfmpc_422_22.

[68] F. M. Rubulotta, G. Ramsay, M. M. Parker, R. P. Dellinger, M. M. Levy, and M. Poeze, “An international survey: Public awareness and perception of sepsis*,” *Crit Care Med*, vol. 37, no. 1, pp. 167–170, Jan. 2009, doi: 10.1097/CCM.0b013e3181926883.

[69] A. M. Wendelboe *et al.*, “Global public awareness of venous thromboembolism,” *Journal of Thrombosis and Haemostasis*, vol. 13, no. 8, pp. 1365–1371, Aug. 2015, doi: 10.1111/jth.13031.

[70] Leaper D., Fry D., and Assadian O., “Perspectives in prevention and treatment of surgical site infection - a narrative review of the literature,” *a Compendium of Clinical Research and Practice*, vol. 25, no. 11, pp. 313–323, Nov. 2013.

[71] J. Fuccella, “Using user centered design methods to create and design usable Web sites,” *In Proceedings of the 15th international conference on computer documentation*, pp. 69–77, 1997.

[72] J. Nielsen, “Users first: How to structure your Website.” Accessed: May 28, 2023. [Online]. Available: www.zdnet.com/devhead/stories/articles/0,4413,2253113,OO.htm

[73] Donald E. Zimmerman and Carol Akerelrea, “A Group Card Sorting Methodology for Developing Informational Web Sites,” *IEEE*, 2002.

[74] J. Mula, R. Barbastefano, M. Díaz-Madroñero, and R. Poler, “Lecture Notes in Management and Industrial Engineering,” 2018. [Online]. Available: http://www.springer.com/series/11786

[75] J. Sageder, A. Demleitner, O. Irlbacher, and R. Wimmer, “Applying Voting Methods in User Research,” in *Proceedings of Mensch und Computer 2019*, New York, NY, USA: ACM, Sep. 2019, pp. 571–575. doi: 10.1145/3340764.3344461.

[76] J. E. Rah, R. P. Manger, A. D. Yock, and G. Y. Kim, “A comparison of two prospective risk analysis methods: Traditional FMEA and a modified healthcare FMEA,” *Med Phys*, vol. 43, no. 12, pp. 6347–6353, Dec. 2016, doi: 10.1118/1.4966129.

[77] “Institutional Review Boards (IRBs) and Protection of Human Subjects in Clinical Trials.” Accessed: Feb. 25, 2023. [Online]. Available: https://www.fda.gov/about-fda/center-drug-evaluation-and-research-cder/institutional-review-boards-irbs-and-protection-human-subjects-clinical-trials

[78] Qualitrics, “Qualitrics.” Accessed: Jun. 20, 2023. [Online]. Available: https://www.qualtrics.com

[79] I. E. Allen and C. A. Seaman, "Likert Scales and Data Analyses," 2007. [Online]. Available: www.sloan-c.org,

[80] F. J. Fowler, "How unclear terms affect survey data.," *Public Opin Q*, vol. 56, no. 2, pp. 218–31, 1992, doi: 10.1086/269312.

[81] Z. M. Alawadi *et al.*, "Facilitators and barriers of implementing enhanced recovery in colorectal surgery at a safety net hospital: A provider and patient perspective," *Surgery*, vol. 159, no. 3, pp. 700–712, Mar. 2016, doi: 10.1016/j.surg.2015.08.025.

[82] J. P. Walker, "Status of the Rural Surgical Workforce," *Surgical Clinics of North America*, vol. 100, no. 5, pp. 869–877, Oct. 2020, doi: 10.1016/j.suc.2020.06.006.

[83] M. Varacallo, T. D. Luo, and N. A. Johanson, *Total Knee Arthroplasty Techniques*. 2022.

[84] American Counts Staff, "MONTANA: 2020 Census," United States Census Bureau.

[85] S. H. Mitchell, K. Hildenbrand, and K. Pietz, "Emergency Physicians' Knowledge of Sports-Related Concussion, Referral Patterns, and Use of Return to Play Guidelines," *Athl Train Sports Health Care*, vol. 8, no. 5, pp. 209–215, Sep. 2016, doi: 10.3928/19425864-20160617-01.

[86] John Rapoport, Daniel Teres, Yonggang Zhao, and Stanley Lemeshow, "Length of Stay Data as a Guide to Hospital Economic Performance for ICU Patients," *Med Care*, vol. 41, no. 3, pp. 386–397, 2003, Accessed: Nov. 08, 2022. [Online]. Available: https://www.jstor.org/stable/3767848?seq=1#metadata_info_tab_contents

[87] J. Wiens and E. S. Shenoy, "Machine Learning for Healthcare: On the Verge of a Major Shift in Healthcare Epidemiology," *Clinical Infectious Diseases*, vol. 66, no. 1, pp. 149–153, Jan. 2018, doi: 10.1093/cid/cix731.

[88] S. Bacchi, Y. Tan, L. Oakden-Rayner, J. Jannes, T. Kleinig, and S. Koblar, "Machine learning in the prediction of medical inpatient length of stay," *Internal Medicine Journal*, vol. 52, no. 2. John Wiley and Sons Inc, pp. 176–185, Feb. 01, 2022. doi: 10.1111/imj.14962.

[89] A. Arora *et al.*, "Predictive Models for Length of Stay and Discharge Disposition in Elective Spine Surgery: Development, Validation, and Comparison to the ACS NSQIP Risk Calculator," *Spine (Phila Pa 1976)*, vol. 48, no. 1, pp. E1–E13, Jan. 2023, doi: 10.1097/BRS.0000000000004490.

[90] C. A. Riley *et al.*, “NSQIP as a Predictor of Length of Stay in Patients Undergoing Free Flap Reconstruction,” *OTO Open*, vol. 1, no. 1, p. 2473974X1668569, Jan. 2017, doi: 10.1177/2473974X16685692.

[91] Jessica R. Burgess, Benjamin Smith, Rebecca Britt, Leonardo Weireter, and Travis Polk, “Predicting Postoperative Complications for Acute Care Surgery Patients Using the ACS NSQIP Surgical Risk Calculator,” *Am Surg*, 2017.

[92] Y. Ma *et al.*, “Assessment of the NSQIP Surgical Risk Calculator in Predicting Microvascular Head and Neck Reconstruction Outcomes,” *Otolaryngology–Head and Neck Surgery*, vol. 160, no. 1, pp. 100–106, Jan. 2019, doi: 10.1177/0194599818789132.

[93] V. Ganganwar, “An overview of classification algorithms for imbalanced datasets,” 2012. [Online]. Available: www.ijetae.com

[94] Ifra Altaf, Muheet Ahmed Butt, and Majid Zaman, “A Pragmatic Comparison of Supervised machine Learning Classifiers for Disease Diagnosis,” *Proceedings of the Third International Conference on Inventive Research in Computing Application*, 2021, Accessed: Feb. 02, 2023. [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9544582

[95] Terry Therneau, Beth Atkinson, and Brian Ripley, “Package ‘rpart.’” Oct. 21, 2022. Accessed: Feb. 02, 2023. [Online]. Available: https://github.com/bethatkinson/rpart, https://cran.r-project.org/package=rpart

[96] T. Whare Wananga, W. Hamilton, N. Zealand, and E. Frank, “mvers1ty ofWaikato Pruning Decision Trees and Lists,” 2000.

[97] J. Machado, F. Soares, J. Trojanowska, and E. Ottaviano, “Lecture Notes in Mechanical Engineering Innovations in Mechanical Engineering.” [Online]. Available: http://www.springer.com/series/11693

[98] T. Chai and R. R. Draxler, “Root mean square error (RMSE) or mean absolute error (MAE)?,” *Geosci. Model Dev. Discuss*, vol. 7, pp. 1525–1534, 2014, doi: 10.5194/gmdd-7-1525-2014.

[99] A. de Myttenaere, B. Golden, B. le Grand, and F. Rossi, “Mean Absolute Percentage Error for regression models,” *Neurocomputing*, vol. 192, pp. 38–48, Jun. 2016, doi: 10.1016/j.neucom.2015.12.114.

[100] “ASA Physical Status Classification System.”

[101] M. A. Coutinho-Myrrha *et al.*, “Duke activity status index em doenças cardiovasculares: Validação de tradução em Português,” *Arq Bras Cardiol*, vol. 102, no. 4, pp. 383–390, 2014, doi: 10.5935/abc.20140031.

[102] K. Lee, H. Kim, H. Lee, and D. Shin, “Flexible Group-Level Pruning of Deep Neural Networks for On-Device Machine Learning,” in *2020 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, IEEE, Mar. 2020, pp. 79–84. doi: 10.23919/DATE48585.2020.9116287.

[103] Stephen Milborrow, “Plotting rpart trees with the rpart.plot package,” Jun. 2021. Accessed: Feb. 11, 2023. [Online]. Available: http://www.milbo.org/rpart-plot/prp.pdf

[104] A. P. Patel, C. Gronbeck, M. Chambers, M. A. Harrington, and M. J. Halawi, “Gender and Total Joint Arthroplasty: Variable Outcomes by Procedure Type,” *Arthroplast Today*, vol. 6, no. 3, pp. 517–520, Sep. 2020, doi: 10.1016/j.artd.2020.06.012.

[105] Samer S. Hasan, Jordan M. Leith, Kevin L. Smith, and Frederick A. Matsen, “The distribution of shoulder replacement among surgeons and hospitals is significantly different than that of hip or knee replacement,” *J Shoulder Elbow Surg*, vol. 12, no. 2, pp. 164–169, 2003.

[106] C. D. Lopez, V. Boddapati, A. L. Neuwirth, R. P. Shah, H. J. Cooper, and J. A. Geller, “Hospital and Surgeon Medicare Reimbursement Trends for Total Joint Arthroplasty,” *Arthroplast Today*, vol. 6, no. 3, pp. 437–444, Sep. 2020, doi: 10.1016/j.artd.2020.04.013.

[107] A. L. Beam and I. S. Kohane, “Big Data and Machine Learning in Health Care,” *JAMA*, vol. 319, no. 13, p. 1317, Apr. 2018, doi: 10.1001/jama.2017.18391.

[108] D. Jin, E. Sergeeva, W. Weng, G. Chauhan, and P. Szolovits, “Explainable deep learning in healthcare: A methodological survey from an attribution view,” *WIREs Mechanisms of Disease*, vol. 14, no. 3, May 2022, doi: 10.1002/wsbm.1548.

[109] K. G. Prasad, B. G. Nelson, C. R. Deig, A. L. Schneider, and M. G. Moore, “ACS NSQIP Risk Calculator,” *Otolaryngology–Head and Neck Surgery*, vol. 155, no. 5, pp. 740–742, Nov. 2016, doi: 10.1177/0194599816655976.

[110] H. Salah, “Predicting Inpatient Length of Stay in Western New York Health Service Area Using Machine Learning Algorithms,” Binghampton University, Binghampton, 2013. Accessed: Aug. 05, 2023. [Online]. Available: https://www.proquest.com/openview/6a5a990e1ec4f2728dfe913a5e47bf3a/1?pq-origsite=gscholar&cbl=18750

[111] America Counts Staff, “What is Rural America?” Accessed: Feb. 12, 2023. [Online]. Available: https://www.census.gov/library/stories/2017/08/rural-america.html

[112] A. I. Edelstein *et al.*, “Can the American College of Surgeons Risk Calculator Predict 30-Day Complications After Knee and Hip Arthroplasty?,” *J Arthroplasty*, vol. 30, no. 9, pp. 5–10, Sep. 2015, doi: 10.1016/j.arth.2015.01.057.

[113] A. D. Blann and G. Y. H. Lip, “Venous thromboembolism,” *BMJ*, vol. 332, no. 7535, pp. 215–219, Jan. 2006, doi: 10.1136/bmj.332.7535.215.

[114] Sujata Dash, Bishwa Ranjan Acharya, Mamta Mittal, Ajith Abraham, and Arpad Kelemen, *Deep Learning Techniques for Biomedical and Health Informatics*, vol. 68. Cham: Springer International Publishing, 2020. doi: 10.1007/978-3-030-33966-1.

[115] J. Morgan, “SAMPLE SIZE AND MODELING ACCURACY OF DECISION TREE BASED DATA MINING TOOLS,” 2003.

[116] R. Sola *et al.*, “Impact of Mentoring on Diversity and Inclusion in Surgery,” *Am Surg*, vol. 87, no. 11, pp. 1739–1745, Nov. 2021, doi: 10.1177/00031348211047486.

[117] Andis Robeznieks, “Native Americans work to grow their own physician workforce,” American Medical Association. Accessed: Jul. 18, 2023. [Online]. Available: https://www.ama-assn.org/delivering-care/health-equity/native-americans-work-grow-their-own-physician-workforce