Creating Operating Room Schedules by Weighted Monte Carlo Simulations

Matthew A. Rock

University of Wisconsin – Green Bay

M.S. in Data Science

April 28, 2023

Abstract

Operating rooms generate over half of hospital revenue, and require careful coordination between physicians, patients, operating room staff and other departments to smoothly and successfully perform each surgery. This paper explores an alternative to the common practice of scheduling case lengths by an average of historical surgery lengths by balancing empty operating room time with case delays. Each scheduled case was simulated with real case lengths corresponding to the physician and billing code selected. The next case is scheduled at the start time that resulted in the lowest total minutes after applying a weighing penalty to the total length of the next case delay and empty operating room time. These schedules were compared to the same five cases scheduled by average case length by the average end time of the final case and total length of all case delays. Ratios where empty operating room time were penalized at one to two times as heavily as case delays performed similarly to scheduling by average case lengths. Linear regression equations based on the comparison metrics for each schedule showed that case delays were inversely correlated with when the final case ends. Another set of simulations that looked at how different orders of the same three cases could affect delays and total operating time found that there was mostly a few minutes saved from the worst to best ordering, with a few schedules showing a substantial reduction of up to 10% of the length of the total operating room time.

Keywords: operating room schedules, Monte Carlo simulations, linear regression, operating room efficiency, surgical delays

Table of Contents

List of Tables 5

List of Figures 5

Chapter 1 – Introduction 8

  Background 8

  Conceptual Framework and Purpose 9

  Study Outline, Significance, and Limitations 9

  Definition and Explanation of Terms 11

    Operating Room Timing Metrics 11

    Overview of Operating Room Resources 11

    Procedure Codes Complexity 12

    Operating Room Scheduling 13

Chapter 2 – Literature Review 13

  Surveys of Operating Room Scheduling Papers 13

  Modeling and Simulations 15

  Other Operating Room Metrics 18

  Impacting Patient Care 20

  Summary 21

Chapter 3 – Data Analysis and Simulation Setup 21

  Procedural Analytics Data Overview and Processing 22

  Grouping Cases 26

Other Timing Intervals 31

An Initial Simulation 34

Improving the Simulation Model 39

Creating and Simulating Operating Room Schedules 41

The Cutting Room Floor 42

Chapter 4 –Results 43

A Complicated Simulation 44

More Weighted Simulations with One Physician and Billing Code 48

Projecting Results by Linear Regression 53

Projecting Linear Regressions by Linear Regression 54

Simulations Using Different Billing Codes 55

Impact of Case Order 60

Chapter 5 - Analysis & Conclusion 63

Scheduling by Weighing Compared to Case Average 63

Case Ordering 64

Future Steps 65

Conclusion 67

Bibliography 68

Appendices 72

Appendix A: Additional Case Re-Ordering Results 72

List of Tables

Table 1 Important Operating Room Timing Metrics 11

Table 2 Original Surgical File Field Descriptions 22

Table 3 Additional Surgery Case Categories 23

Table 4 Provider's Physicians and Case Information 24

Table 5 Shapiro-Wilk Normality Test Results for Surgical Case Lengths 26

Table 6 Skewness of Physician Case Lengths by Times Performed 30

Table 7 Distribution of Wheels In to Cut Case Lengths by Service Line, Ordered By Service Line Case Count 31

Table 8 Distribution of Close to Wheels Out Case Lengths by Service Line, Ordered By Service Line Case Count 32

Table 9 Average Turnaround Time based on Case Length and Prior Surgeon 33

Table 10 Confusion Matrix of Simulated Cases Starting on Time or Late 35

Table 11 Cases Being On-Time or Late As Result of Prior Case 36

Table 12 Case Delays by Scheduled Hour of Day 39

Table 13 Final Data Set Overview 42

Table 14 Time Allotted for Total Knee Replacement Cases in Schedules by Ratio 47

Table 15 Scheduled Case Lengths as Percentile of Cases in Billing Code Simulations 52

Table 16 Results of Linear Regressions on Prior Simulation Linear Regressions 55

Table 17 Multiple Billing Code Simulation Overview 55

List of Figures

Figure 1 P-Values of Laparoscopic Appendectomies 27

Figure 2 P-Values of Total Knee Arthroscopies 27

Figure 3 P-Values of Coronary Artery Bypass Grafts 28

Figure 4 Kruskal-Wallis Provider - Billing Code P Values by Case Count 29

Figure 5 Levene Provider - Billing Code P Values by Case Count 29

Figure 6 Simulated Cases Timeliness by Order in Day 35

Figure 7 Histogram of Case Lengths Used in Simulation.. 38

Figure 8 Average End of Simulated Knee Surgeries by Weighted Scheduling... 45

Figure 9 Average Total Delays of Simulated Knee Surgeries by Weighted Scheduling............................... 46

Figure 10 Weighted Simulated Total Knee Replacement Case Timeliness with Ratio 1.6 By Order In Day 47

Figure 11 Mean and Variance for Additional Weighted Simulation Physician-Billing Codes 48

Figure 12 All Single Billing Code Simulation Results Compared to Baseline 49

Figure 13 Average of Length of Operating Day by Simulation over Baseline for Single Case Simulations.. 50

Figure 14 Best Performing End Times Compared to Baseline, All Single Billing Code Simulations............ 51

Figure 15 Average of Total Case Delays over Baseline, All Single Billing Code Simulations 52

Figure 16 Coefficient of Determinations from Linear Regression Modeling of Metrics Based on Ratio 54

Figure 17 All Multiple Billing Code Case Simulation Results Compared to Baseline.................................. 56

Figure 18 Ratio of Total Case Delays over Baseline, All Multiple Billing Code Simulations by Ratio........... 57

Figure 19 Ratio of Final Case End Times Compared to Baseline, All Multiple Billing Code Simulations by Ratio 58

Figure 20 Coefficient of Determination for Multiple Billing Code Linear Regressions............................... 59

Figure 21 Best Performing End Times Compared to Baseline, All Multiple Billing Code Simulations 60

Figure 22 Percent Difference of Shortest and Longest Average End Times for Reordered Cases Over Longest Avg End by Ratio 61

Figure 23 Percent Difference of Least and Most Case Delays for Reordered Cases Over Most Case Delays by Ratio 62

Figure 24 Three Reordered Cases by Total Delay and Avg Final Case End with Five Different Scheduling Methods..63

Chapter 1 – Introduction

Background

The COVID-19 pandemic has disrupted seemingly every single part of the economy. Those disruptions naturally started with the health care sector, where hospitals cancelled or postponed elective surgeries to free inpatient beds for those suffering from the new disease. The underlying causes for those surgeries were untreated, while other candidates for surgery accumulated as well, creating a substantial backlog of cases. One estimate of the time to get the United Kingdom's waiting list for hip and knee arthroplasty cases to a pre-pandemic level is between 20 and 48 months of additional 30% capacity (Oussedik, et al., 2021). Major hospitals in Washington and Indiana had to postpone thousands of cases (Advisory Board, 2022).

As critical as performing a timely surgery is to the health of a patient, it is just as important to the financial health of a hospital. The operating room generates 65% of the revenue of a hospital (Peters, Basham, & Cole, 2021). A minute of operating room time can cost between $30 to $100 (CaseCtrl, 2021). Maintaining improvements to scheduling that clears the COVID-19-induced backlog of surgeries will help financially sustain hospitals that provide critical services to their communities.

Each resource required to perform one surgery is very limited. Becoming a surgeon takes years of medical school and residency training, so there's only so many qualified individuals that can perform a given operation. Health care staff turnover has dramatically risen in the wake of the pandemic, "with hospital turnover exceeding every previous survey conducted by NSI Nursing Solutions Inc. During the past year, hospital turnover increased by 6.4% and currently stands at 25.9%" (Colosi, 2022). The physical operating room space is limited as well, as building even a single ambulatory surgery center operating room costs an average of $1.4 million (Hatton, 2022).

The purpose of this research is to identify strategies to schedule surgical cases in an operating room by using Monte Carlo simulations to adjust when cases start based on weighing minutes of case delays and empty operating room differently. This method will allow to schedule based on reducing staff work hours or minimizing case delays, compared to the standard method of scheduling by case duration of the average of the length of the surgeon performing similar cases (Common Ground Editorial Board, 2021).

Conceptual Framework and Purpose

As an example of the difficulty in operating room scheduling, imagine an extremely bimodal process that takes one hour half of the time, and two hours other times, but is impossible to know how long each specific execution of this process would take beforehand. Would it make the most sense to schedule this process for one hour, two hours, or ninety minutes? Should every process scheduled have the same length of time, or should consideration be given to what could happen in prior events? Each approach has strengths and drawbacks and would depend on balancing the need to have an accurate schedule that minimizes delays with the penalties of assigning tasks that run longer than the planned allotment of time.

This research will explore how the timing and ordering of scheduling surgical cases can affect the operating room's efficiency by answering the following questions:

1. How can operating room schedules balance the need to reduce downtime between cases with patient dissatisfaction from having their surgeries delayed?
2. Does scheduling cases in a certain order impact the length of time it takes to complete the cases or how long cases are delayed?

Study Outline, Significance, and Limitations

The study will use surgical data from 120 hospitals to form Monte Carlo simulations of operating room schedules to analyze how to schedule cases to minimize delays and maximize operating room efficiency. The scheduling model developed will be compared to schedules built on average case lengths.

Current operating room scheduling focuses on individual cases. Each individual case's length is determined by either the surgeon's estimate or an average of the recent times the case was performed. Differences in the scheduled case length to the actual length of each case can affect the rest of the day in the operating room. If a case runs longer than scheduled, each subsequent case may be delayed and could lead to unexpected overtime for staff. But extra time between cases can lead to unused operating room time, and extra minutes here or there could have added up to another case. This research will create strategies for scheduling cases based on not just historical case lengths, but the prior cases performed in the day, helping hospital operating room management reduce the downtime between cases and staff overtime.

This research will be novel because of the uniqueness of the Vizient® Procedural Analytics data set (Vizient, Inc., 2023). Vizient Procedural Analytics is a data-focused software platform containing surgical case data submitted by participating health care providers and is used to compare costs and quality outcomes based on types of surgeries. Existing studies on operating room efficiency look at a single hospital or system of hospitals. This data set contains key operating room times and primary billing code information for operating rooms in 120 different hospitals. Scheduling methods can be tested and validated at a variety of locations with their own surgeons, case mix, and protocols. This makes potential strategies of interest relevant to a wider variety of hospitals, as a single hospital's findings may be characterized as only applicable to that single hospital's characteristics (Slyter, 2018).

The wide variety of hospitals, physicians, and types of surgeries is balanced by a narrow focus of included data fields, leading to the limitations of this study. Complicated cases where one surgery had multiple procedures performed will look like only one procedure. A hospital with a very short case setup could be doing the work while the patient is out of the room. A patient could be in the operating room waiting for space in the recovery unit. All these possibilities are not available to be explored in this data set.

Definition and Explanation of Terms

Operating Room Timing Metrics

Table 1 Important Operating Room Timing Metrics

Metric	Description
First Case On-Time Start	Percentage of first cases in an operating room where the patient entered the room at or before the scheduled start time divided by the total amount of first cases.
Wheels in to cut	Time in minutes the patient is in the operating room before the surgeon makes the initial incision. Includes patient positioning, induction of anesthesia, and equipment setup.
Cut to close	Time in minutes where the patient's skin has been cut to facilitate the surgery to when all incisions are closed.
Close to wheels out	Time in minutes between the closing of the incision and when then patient leaves the operating room to the recovery area.
Wheels in to wheels out	Time in minutes the patient is in the operating room.
Turnaround Time	Time in minutes between the end of one case and the start of the next case in the same room. OR staff "turn" the room over by removing used equipment, cleaning, and setting up the next case.
Case delay	Time in minutes between the scheduled case start and the patient enters the room.
OR utilization	Total wheels in to wheels out time during staffed operating room hours, divided by total staffed operating room hours (Ayres & Trout, 2022).

Overview of Operating Room Resources

There are several key people and resources that are required to perform surgery. In non-emergency cases, the patient to be operated on needs to be prepared before being wheeled into the operating room. The pre-operation routine includes being declared healthy enough for surgery to be performed and consenting to the operation in writing. The surgeon needs to be ready and available as well, having completed any prior cases on their schedule or other hospital duties. There must be a small clinical team of operating room nurses available to assist the surgeon during the operation, as well as an open and clean operating room. In the time between the patient entering the operating room and the initial incision, an anesthesiologist must properly sedate the patient and surgical tools must be available in the sterile field.

Procedure Codes Complexity

To standardize the diagnosis of a patient, the method of treatment, and appropriately billing for the severity and complexity of care, the health care industry uses several types of medical coding. The two that involve surgery are the International Classification of Diseases, Tenth Revision, Procedure Coding System (ICD-10) and Current Procedural Terminology (CPT). (Jones, 2018) CPT codes use a five-digit number to describe the treatment, with similar procedures grouped by their number. ICD-10 codes use seven alphanumeric characters, with each character serving to identify the location on the body, root operation, approach, and if a device or implant was critical in the patient care. Surgeries in an outpatient setting bill using CPT codes, and inpatient surgeries will use ICD-10 codes to bill the patient, while the physician will bill using CPT codes.

To illustrate the complexity of coding and surgery, consider an appendectomy. There are four different CPT codes and seventeen ICD-10 codes that indicate an appendectomy. The four CPT codes are for an appendectomy with an open approach, a laparoscopic appendectomy, an appendectomy where the appendix ruptured, and an appendectomy that was performed in addition to another procedure. The different ICD-10 codes include how the appendix was removed, what the opening was, and if the process was diagnostic or not.

Operating Room Scheduling

Creating a day's operating room schedule takes place well before that day arrives. Operating room managers will assign specific days to a surgeon, service line or surgical group for the foreseeable future in a process called block scheduling. (Sadler, 2016) Surgeons value the ability to know they have dedicated time to schedule cases in advance, and hospitals offer blocks to surgeons to entice them to bring patients to their location over others. Other resources, like a robotic surgical system or multiple rooms assigned to one surgeon to reduce their surgeon turnaround times, may also be made available depending on hospital resources.

Chapter 2 – Literature Review

With the amount of schooling required to become a practicing physician and the number of hospitals affiliated with a medical school or larger university structure, it should come as no surprise how many papers are published studying an operating room's organization. The importance of the operating room in the hospital, the complexity of aligning resources, and the impact on patient care means many fields relating to health care and computer modeling have studied this problem.

Surveys of Operating Room Scheduling Papers

In their review of operating review planning, Zhu et al. (2019) divided their papers into six categories: decision levels, scheduling strategies, patient characteristics, problem features, mathematical models, and solutions and methods (Zhu, Fan, Yang, Pei, & Pardalos, 2019). The three decision levels describe the amount of time and resources in scope, with the strategic level involving long-term surgical specialty recruitment, facility planning, and capacity allocation, the tactical level focusing on allocating operating room time by specialty every month or few months, and the operational level of scheduling cases by advance (an OR and day) or allocation (the procedure start time). Scheduling strategy papers detailed using block scheduling, open scheduling, or a combination of the two. Patient characteristic papers focused on inpatient versus outpatients and elective or non-elective patients. The problem features category focuses on of the study like variance (including case length variance) and clinical requirements like room cleaning. Finally, they found multiple kinds of models and algorithms studied to provide solutions to the problem of operating room scheduling.

Another overview by Rahimi and Gandomi (2020) identified many of the same patterns with their focus on research papers with algorithmic descriptions (Rahimi & Gandomi, 2020). They included more metacontextual analysis in their paper, especially regarding the number of published articles before 2005, from 2005 to 2010, and 2010 to 2019. Significantly more articles were published on solutions using waiting time, overtime, and utilization as the model's criteria. One of their final points is that "there is limited software available as user-interference applications for these specific problems, and there is a critical need for more investigation in this field." A similar survey by Guerriero and Guido (2011) on operational research papers identified the same strategic levels, uncertainty characteristics, and models (Guerriero & Guido, 2011). They noticed the operational level papers focused on "patient satisfaction and resource efficiency maximization, measured in terms of patient waiting time, OR utilization and costs."

Taken together, the three surveys show that there is a tremendous amount of variation in approaches for operating room efficiency research. Even the idea of "operating room utilization" would be answered differently by a CEO's long-term view compared to the operating room manager figuring out how to adjust cases for a key staff member calling out sick. The variation in limiting criteria, variables to maximize, and mathematical approaches is the result of the different problems a hospital faces, and the differing skill sets of analytical capabilities available. There were so many combinations of approaches that while each survey specifically mentions Monte Carlo simulations, none of the papers using the technique were at the allocation level of scheduling this study focused on.

Modeling and Simulations

A recent paper from Abbou et al. (2022) showed the capabilities and limitations of machine learning in predicting surgical case length (Abbou, Tal, Frenkel, Rubin, & Rappoport, 2022). Their XGBoost model was trained on 254,623 surgeries performed at two hospitals over a ten-year timeframe and used an extensive amount of preoperative data gathered, like drugs administered and the surgeon's experience. It was then compared to a naïve model of simply using the median case length for all surgeries in the training data having the same set of procedures. Both predictions performed similarly overall. The XGBoost model was more accurate at predicting surgical case length at the hospital with a larger proportion of shorter cases, while it did about as well as the naïve model at the hospital with longer cases. There was a correlation between the median length of a surgery with both the mean absolute error and root mean squared error. Critically, the paper didn't show enough granular data presented to indicate exactly when the complex, data-intensive model would be a better use case than a simple average, just a very high-level summary of error summaries that showed small differences. An analysis of when the complex model outperformed using the average case time would have provided a better starting point for further research as well as for the hospital's own reference in future scheduling.

Another look at variations on that naïve model by Wu, Huang et al. also left gaps in the analysis (Wu, Huang, Weaver, & Urman, 2016). They compared a rolling average of the case length of the last 3, 5, 10 and 20 total knee arthroplasties to the surgeon's prediction of the length of that case. The surgeon's average resulted in an average of 18.1 minutes of overbooking, while the rolling averages mean difference to the actual case length was under a minute. Pointing out that the average difference between each case length and the rolling average of recent case lengths is close to zero has limited application. Most of their analysis focused on the possible time savings from being more accurate and not overbooking but ignored the variance in their estimates. Not knowing which cases will run over or under their estimates by the standard deviation of 16 minutes means later cases will be delayed longer and more frequently. Additionally, their results may be biased due to the three surgeons whose predictions were not statistically significantly different than the rolling average having only performed 514 of the 2539 analyzed cases.

Looking at the bigger picture of a day in an operating room and the cases scheduled, a machine learning model dramatically improved the number of rooms that finished within fifteen minutes of the scheduled end time (Rozario & Rozario, 2020). By applying their model to scheduling cases, the rooms finishing on-time increased from 15% of the time to 55%, while completing 97% of the previous volume of cases with the same number of OR minutes used. The reduction in revenue-generating cases may be an issue but is balanced by a projected savings of $469,000 in overtime pay reduction over three years. As the survey of models underscored, hospitals have different problems. If staff retention is an issue, ensuring they can go home on a regular schedule helps prevent burnout and keeps them from working elsewhere.

A similar look at rearranging scheduling times using a branch-and-bound algorithm led to savings in staffing costs (Li, Gupta, & Potthoff, 2016). After confirming the hospital's sampling of projected case lengths was a reasonable measure for the actual case lengths, and that case ordering has no discernable effect on the lengths, Li, et al. rearranged the scheduled cases into different operating rooms. They leveled the number of operating rooms used throughout the day to limit a dip from the morning to the afternoon, reducing the amount of total staffed rooms required. However, some of those staffing savings would result in overtime pay at one of the three hospitals, due to their longer-than-average case times.

The authors mentioned the algorithm's increase in surgeon idle time as another concern. The algorithm could place surgeons with multiple cases in a way that they'll start the day, wait for another surgeon to complete their work, and then come back to the operating room and end the day. The detailed algorithmic notations by Li, et al. that show how this issue arises is a sharp contrast to Rozario and Rozario's work, which was not included in their published paper but referred to as Appendix A via a URL that no longer works.

A team at Massachusetts General Hospital went beyond the computer simulation and piloted a change in their workflow (Sokal, Craft, Chang, Sandberg, & Berger, 2006). They used a new approach to induce anesthesia for surgical patients to inform their modeling. Their parallel processing model used a separate induction room to start anesthesia for the next patient while their operating room was turned over from the previous surgery, rather than waiting for room prep to finish, wheeling in the patient, and then inducing anesthesia. After applying their findings from the parallel processing trial to 49,887 sequentially performed cases, 26% yielded an improvement in operating room time large enough to add another surgery in their nine hour OR day. However, the full improvements in efficiency couldn't be realized. Due to limited space in their hospital's post-surgery recovery room, they also looked at parallel processing using three operating rooms with a fourth as a mini-recovery room. Even with removing an entire room, a wing of three parallel processing rooms and mini-recovery room outperformed four serial processing rooms.

It's no coincidence that the workflow change at Massachusetts General was the model that demonstrated the largest possible improvement in performance. Each model that was strictly computer-based was limited by established processes in the real world, from the inherent variance in surgery length limiting the estimation improvement in a machine learning model to balancing surgeon preference with hospital priorities for daily scheduling. Those computer models must be based on what has happened in the past to be accepted as a realistic possibility. By changing their anesthesia process in real life, the Massachusetts General team's simulations applied their trial run to a larger data set. There is a big advantage in being able to simulate the impact of a process. A computer model built only on the assumption that anesthesia prep can begin in another room wouldn't have any data on the challenges of patient movement or timing between each step. When an overly optimistic result can't be replicated in real life, the entire project could be questioned.

Another difference in the models is the goals. The takeaway point in each abstract mentioned overtime staffing costs, operating room throughput, decreased recovery room workload, and surgical case length variance, again showing the diversity of operating room scheduling issues at different hospitals.

Other Operating Room Metrics

There are a few additional measurements that operating room managers use that are important to consider. One common metric is first case on-time starts. The first scheduled case of the day in a room is prized by surgeons since there's no possibility of an earlier case running late. That same logic is why they're monitored – delays at the start of a day not only mean an empty operating room but put the rest of the day's cases behind schedule. One retrospective study at a hospital of 17 staffed operating rooms found that of the 3604 first cases performed, 55% were delayed, with half of the reasons for delay being an unavailable patient or surgeon (Hicks, Glaser, Scott, Sparks, & McHenry, 2020). They estimated the total delays of 631 hours resulted in an estimated loss of almost $400,000 in idle labor and staff overtime.

Another team looked at the effect of implementing a new electronic medical record system on first case starts (Wu, Kodali, Flanagan, & Urman, 2017). After the first month's deployment issues were resolved, the first case delays were similar to the prior system's first case delays. The time between wheels in to cut declined in average length compared to the old system, leading to a potential savings of $160 per case. Interestingly, their "scheduled" start of 7:30 AM for the first case of the day was acceptable if the patient was in the room at 7:35 AM, showing that even something as seemingly straightforward as when a case is late varies by hospital.

A seven-year retrospective study at one hospital reviewed non-scheduling factors that contributed to all case delays, not just the first case (Pappada, et al., 2022). Their study showed variation in delays by specialty, and increased delays if a preadmission test was required or the patient was significantly overweight. The authors actively used their findings to improve their processes on an annual basis and reduced their average delays by almost 20 minutes per instance by the end of their study. All three papers mention that case delays can cause surgeries to be postponed and can impact patient satisfaction.

Pappada et al. briefly discuss turnaround time, the other major metric. The amount of time between cases is highly dependent on the workflow of individual hospitals. One hospital selected two general surgeons that regularly performed multiple cases daily and reduced their turnaround times from 37 minutes to 14 minutes, as well as inspiring another surgeon to improve his average without the focus of their improvement team (Cerfolio, et al., 2019). Another took that process a step further and reworked their process across their entire operating room team (Fullerton, et al., 2017). Their efforts reduce turnaround times from 45 minutes to 21.7 minutes. Both papers extensively detailed their supporting teams and the cooperation they had in achieving their results.

These two operating room schedule metrics show different underlying priorities. There is one first case in a room a day, while that same room has a turnaround time measurement between every case. The two hospitals that improved their turnover times could potentially add another surgery. Comparatively, the median first case delays were 12 minutes and 6 minutes in the two first case studies respectively. But the first case on-time start is highly visible, even if it's not tracked. The staff is ready to go, and when the clock ticks a minute past it's noticed. It's an easy calculation to make as well as an intuitive one. Tracking those late cases is less about efficiency and more about the politics of an operating room.

Case delays and turnaround time are part of the framework of the operating room schedule, and both directly impact the efficiency of the operating room. Room turnover means that a case can't start, and case delays could lead to staff overtime and unhappy patients.

Impacting Patient Care

A meta-analysis by Cheng et al. reviewed papers that linked the length of a particular surgery with complications (Cheng, et al., 2018). Studies looked at specific cutoff times (comparing surgeries over or under two hours) as well as incremental increases in case length. Of the 66 papers they reviewed, 80% reported a statistically significant association between longer operative times and surgical complications.

Two papers looked at the timing of cases as it impacts the health of a patient. At University of Virginia Health System, a study of 3,395 non-emergent cardiac operations showed that the 10% of cases that started after 3 PM had a higher mortality rate as well as a higher cost of care (Yount, et al., 2015). McIsaac et al. looked at the time between an emergency patient's booking time and entry into an operating room as compared to standard wait time for the patient's priority level (McIsaac, et al., 2017). They found a similar pattern, where patients delayed longer had significantly higher mortality rates, longer lengths of stay, and higher hospital costs.

It is satisfying to see that the goals of the operating room as a business align with the goals of the operating room as a purpose. Hospital initiatives, especially cost saving initiatives, are often weighed against the risk to patient care. Finding ways to better use existing resources while improving patient care helps satisfy two of the "Quadruple Aims" of healthcare (Haverfield, et al., 2020). There may be individual patient concerns, like scheduling patients requiring overnight monitoring later in the day or scheduling diabetic patients earlier to help better control their glucose levels while fasting (The

Cleveland Clinic Foundation, 2018). But as Li, et al. point out, case order has no discernable effect on case length. Fixing one case to a particular slot in the day is easier with more accurate scheduling.

Summary

Improving operating room efficiency is a universal goal with numerous attempts to crack it. Computer modeling can function as a proof-of-concept but needs to be grounded with a clear enough structure to show how and why there is improvement. Details on the variables and models used not only made a particular paper's findings more believable, but seemingly repeatable at other operating rooms. Those papers also tended to look at the impact of their findings on the rest of the operating room, instead of solving their specific problem by creating others.

When focusing on the operational levels of operating room utilization, the terminology held constant. There was a consensus around definitions like efficiency as a percentage of operating time over total staffed hours, turnaround time referring to the room time between cases instead of the surgeon's time, and block scheduling by surgeon or service line. Little differences like considering a case starting at 7:35 for a scheduled 7:30 start on-time may be technically different, but still within the spirit of the nomenclature. There is wide base of established operating room utilization research that provides enough foundation to build upon without any single area established well enough to tower over the rest.

Chapter 3 – Data Analysis and Simulation Setup

Because of the uniqueness and complexity of the Procedural Analytics data set, spending time exploring the data is well served. This chapter starts with an overview of the elements of the data set, looking at ways to separate cases by billing code, physician, and hospital. Then, the results of a simple simulation are analyzed to see empty operating room time and late cases. Those results are used to add complexity to the simulations that will create the different operating room schedules as well as simulated operating room days based on those schedules. Finally, to improve the simulation analysis, cases that would not be helpful are removed from the data set as well as complications that would muddle findings from studying operating room scheduling.

Procedural Analytics Data Overview and Processing

Procedural Analytics (PA) data underlying this study contains details on surgical cases, physicians, and hospitals (see Table 2). The surgical data is submitted regularly by hospitals using the PA platform, so correlations in timing and errors are to be expected on the hospital level. A flat text file containing de-identified data representing 2,936,178 cases was obtained from Vizient. The hospital, physician, and operating room had been blinded to protect patient information, and each individual date of service has been changed. The primary CPT or ICD10 procedure code is unchanged, as are the timestamps for the patient entering and leaving the room as well as the incision cut and close.

A Python script was created to process, format, and categorize the data. The original file contained the following fields.

Table 2 Original Surgical Data File Field Descriptions

Field	Data Type	Description
DE_ID_provider	Categorical text	Identifies hospital or hospital group surgery was performed at.
DE_ID_physician	Categorical text	Identifies physician performing surgery.
DE_ID_suite	Categorical text	Identifies room surgery occurred in. May be null.
CPTICD10Code	Categorical text	Main billing code for surgery.
Date_of_service	Date	When surgery took place.
Schs_time	Time	When surgery was scheduled
Prms	Time	When patient entered surgery room
Inc	Time	When incision was made
Ince	Time	When incision was closed
Prme	Time	When patient left surgery room
Sche	Time	Scheduled end of surgery. May be null.

The Python script combined the date of service with each timestamp, then calculated the minutes between each interval as described in Table 1 Important Operating Room Timing Metrics. Additional categorical fields were added to each case.

Table 3 Additional Surgery Case Categories

Field	Data Type	Description
Scheduled Case Gap	Numeric	Time in minutes between case and the next case in the same room.
Case Order In Room	Numeric	Assigns order to the case based on the wheels in timestamp for that day for all cases in the same room, starting from midnight until 11:59 PM.
Surgeon Case Order	Numeric	Assigns order to the case based on the wheels in timestamp for that day for all cases by the same surgeon, starting from midnight until 11:59 PM.
Surgeon Follows Self	Boolean	TRUE if the preceding case in this case's room was performed by the same surgeon on the same day, FALSE otherwise.

Each included case has the proper ordering of the timestamp sequences, meaning every wheels in to cut, cut to close, and close to wheels out time interval is at least one minute long. No seconds were included, so all timing intervals are calculated in whole minutes. With the longest case being 1089 minutes, which is under the total minutes of a day, and no cases with a negative time interval, it follows that there is no case in the data that lasted overnight. With the data set including a limited number of fields, it is hard to incorporate additional information to justify including or excluding outliers in timing data. If a surgery's incision close to wheels out time is long, it could be because someone didn't record the patient leaving when it happened, and only noticed the issue an hour later, or the incision close time was recorded too early when a check of instruments came up short, or the patient had to stay in the operating room because the post-operation recovery room was at capacity.

The actual procedural suite is a key field to see when one case followed another, but its usefulness is limited. Due to data warehousing priorities, the room the surgery took place in was only recently added to the Procedural Analytics data mart. When the room value is provided, the value may represent the operating room suite, instead of a specific room, giving the appearance in this blinded data that multiple surgeries were simultaneously performed in the same room. Any turnaround time analysis will be limited to hospitals with a significant amount of non-overlapping cases, as the entries in Table 3 would have incorrect values for those cases. Three flags were added to individual cases to find instances where a case has a usable operating room, as follows:

1. The case did not overlap with another case in the same room on the same day. This may occur even in operating room data where the room was provided. Overlapping cases is a clear sign of bad data, but it would be impossible to tell which case or which field has the issue. It's possible that both cases were recorded correctly, but the wrong room was recorded.
2. There was at least one case in the room that day. There may be times to select cases that were the only one performed in a day, but there would be no point in relating that single case to others.
3. The case was performed in the same order it was scheduled in that room. Changes in the schedule means the plan went awry.

If one case failed any of those checks, all other cases performed in that room on that day were given a FALSE flag in the "ORDayCheck" to indicate there may have been data errors elsewhere that may affect the results. While this may seem strict, a third of all cases were on an acceptable operating room day.

Looking for trends in the different elements of data, there is not much that can be considered evenly distributed. The highest level is the provider, representing a hospital or system of hospitals. Each one serves different needs, ranging from ensuring health care access for rural populations to complex academic medical centers in major cities. That diversity is represented in the PA data set.

Table 4 Provider's Physicians and Case Information

Volume of Cases Performed	Number of Providers in Grouping	Avg Physician Count	Avg Count of Distinct Billing Codes
Over 75,000	8	971.6	5,984
Between 50,000 – 750,000	7	487.9	4,648

Between 25,000 – 50,000	24	416	3,108
Between 10,000 – 25,000	30	172	1,859
Under 10,000	52	75	597

Data from Vizient Procedural Analytics used with permission of Vizient, Inc. All rights reserved.

Different physicians' workload followed a similar pattern. The "average" physician performed 94.3 cases over the 730 different days in the study, but 14,367 (47.5%) performed five or less cases, and 281 performed over 1,000. The busiest physicians had a few recurring types of cases: colonoscopies and endoscopies, cataract removals, and epidurals. There was little connecting the types of cases performed by lower-volume physicians, possibly because they only performed one case at the hospital due to extenuating circumstances or because they started or ended their career at the edge of the data set's timeframe. Each physician performed cases at only one provider.

There is also an unbalanced distribution regarding how many times a particular surgery was performed. Of the over 21,000 different primary surgical codes used, 41 were performed over 10,000 times and about 5,700 were used once. Half of the data is included in physician-billing code combinations that occurred at least 20 times. But as the laparoscopic appendectomy discussion above showed, there are numerous billing codes that indicate similar procedures.

Finally, there is some self-selection in the hospitals submitting data to the PA platform. Hospitals would need adequate analytical and financial resources to submit data to PA, and interest in use of the platform. It is important to point out that much of PA's focus is on the cost and quantity of supplies used in the operating room, not case timing. Whatever reasons facilities would have to invest in using PA, they would be primarily based on supply costs, so the case lengths would be minimally affected by the selection bias. The additional resources that allow a hospital to subscribe to PA would suggest other resources like additional staff or expensive equipment like surgical robots.

Grouping Cases

The location, physician and type of procedure are known at the time of scheduling the case. Identifying what populations of case lengths should be grouped together is critical in improving scheduling, so using those three data elements to determine what is meaningfully different enough to be separated to improve accuracy is a logical starting point before combining similar procedures. The first step is testing for normal distributions using Shapiro-Wilk on each combination.

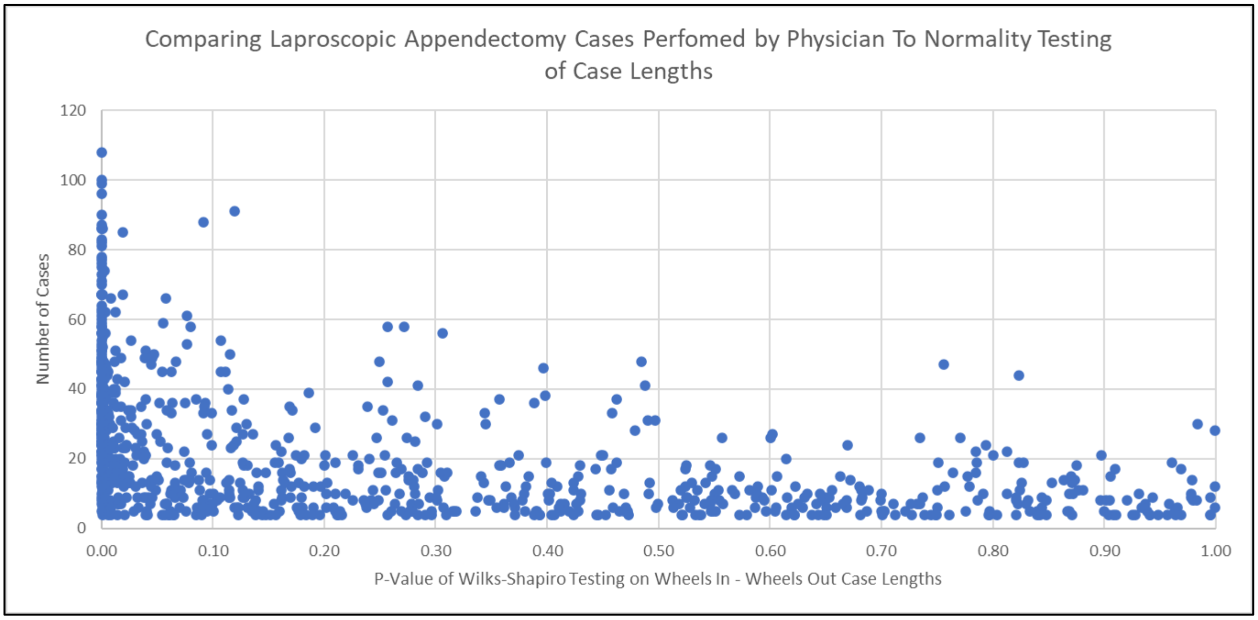
Table 5 Shapiro-Wilk Normality Test Results for Surgical Case Lengths

P Value of Shapiro-Wilk Test	By Procedure Only	By Provider and Procedure	By Physician and Procedure
Under .05	54.3%	34.5%	24.4%
Between .05 and .5	28.6%	38.6%	43.0%
Over .5	17.1%	26.9%	32.5%

Data from Vizient Procedural Analytics used with permission of Vizient, Inc. All rights reserved.

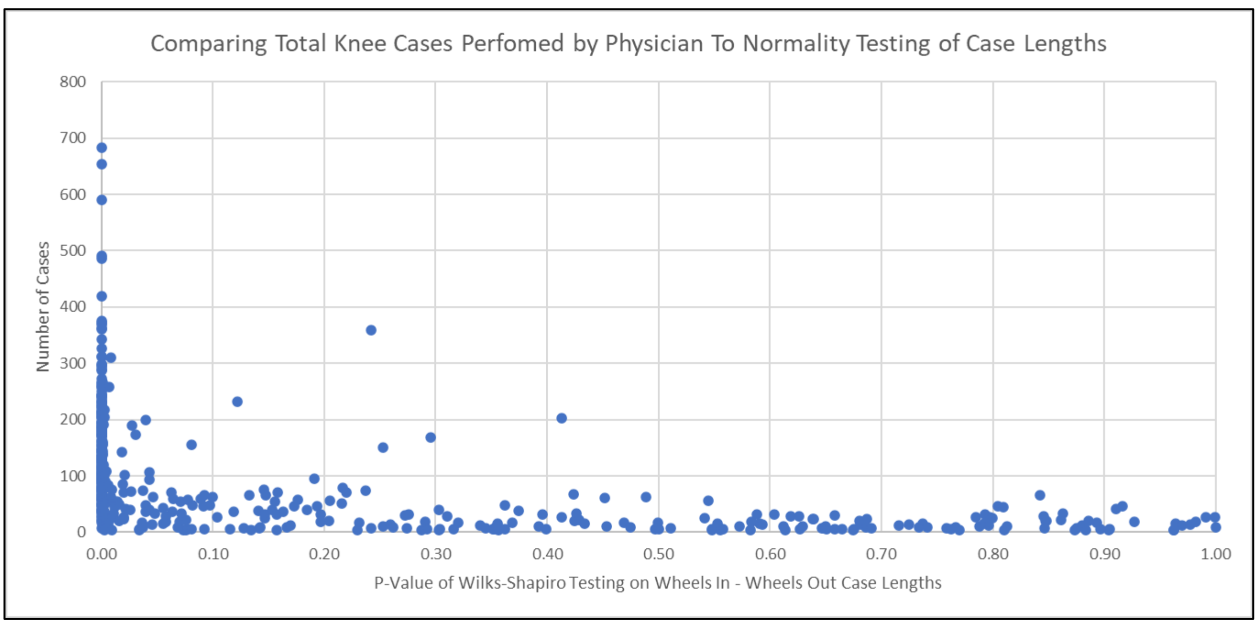
More groupings of case lengths skewed closer to a normal distribution when going by physician and procedure, but that could be as much a byproduct of reducing the number of data points included in each population as it is describing the underlying data. Looking at three common procedures closely helps clarify what level of normality can be expected.

Figure 1 P-Values of Laparoscopic Appendectomies



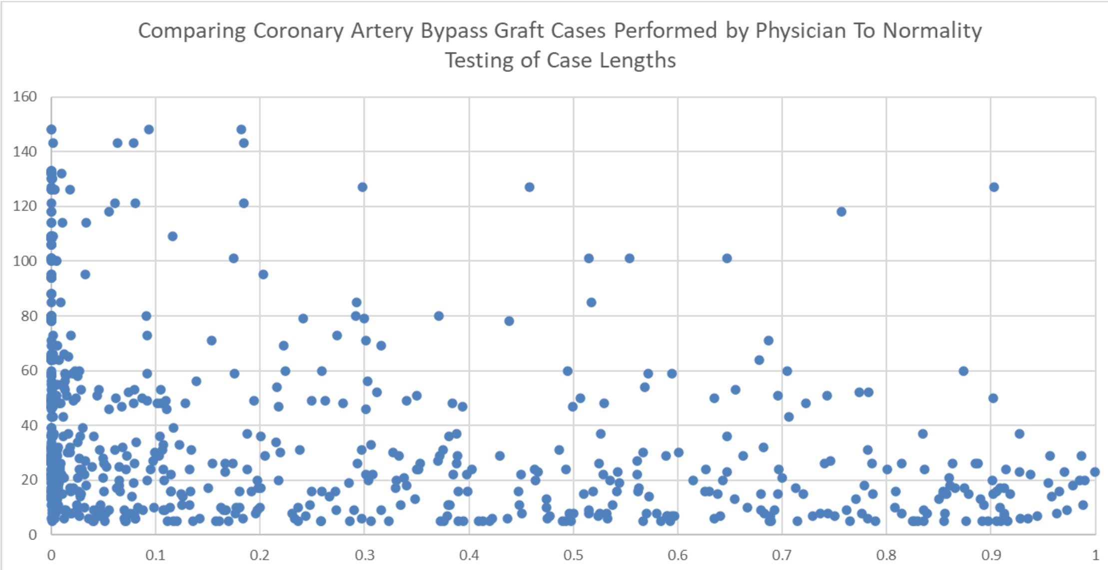
Data from Vizient Procedural Analytics used with permission of Vizient, Inc. All rights reserved.

Figure 2 P-Values of Total Knee Arthroscopies



Data from Vizient Procedural Analytics used with permission of Vizient, Inc. All rights reserved.

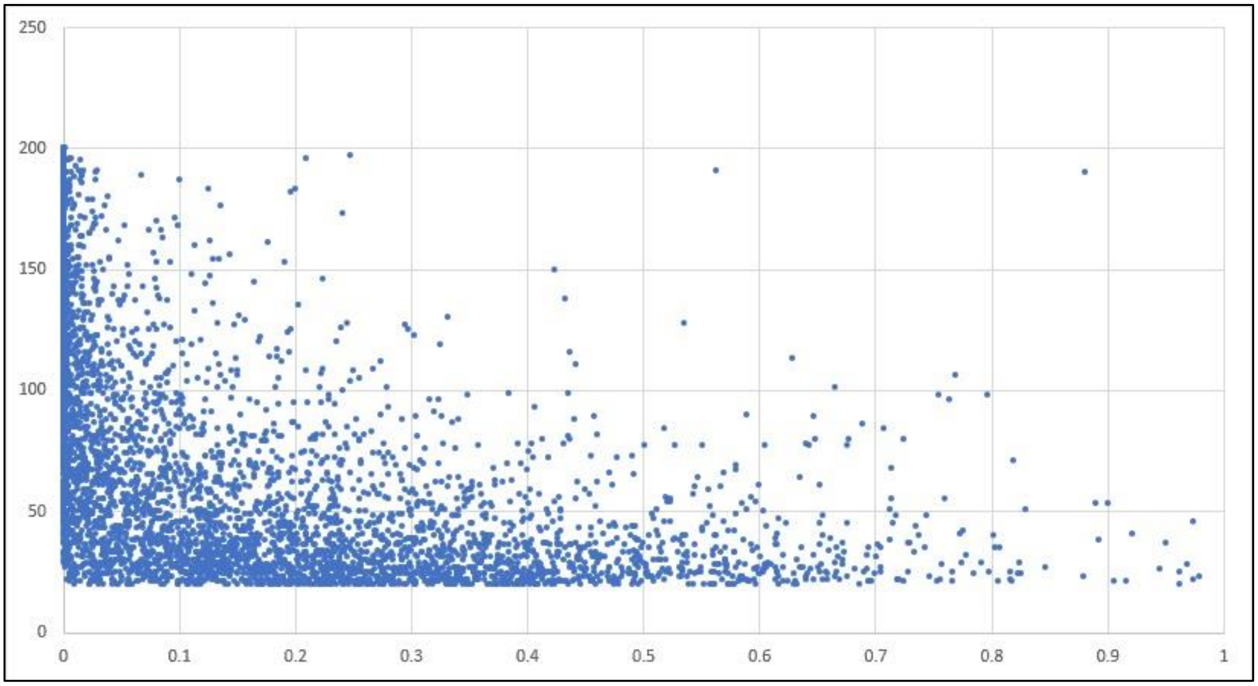
Figure 3 P-Values of Coronary Artery Bypass Grafts



Data from Vizient Procedural Analytics used with permission of Vizient, Inc. All rights reserved.

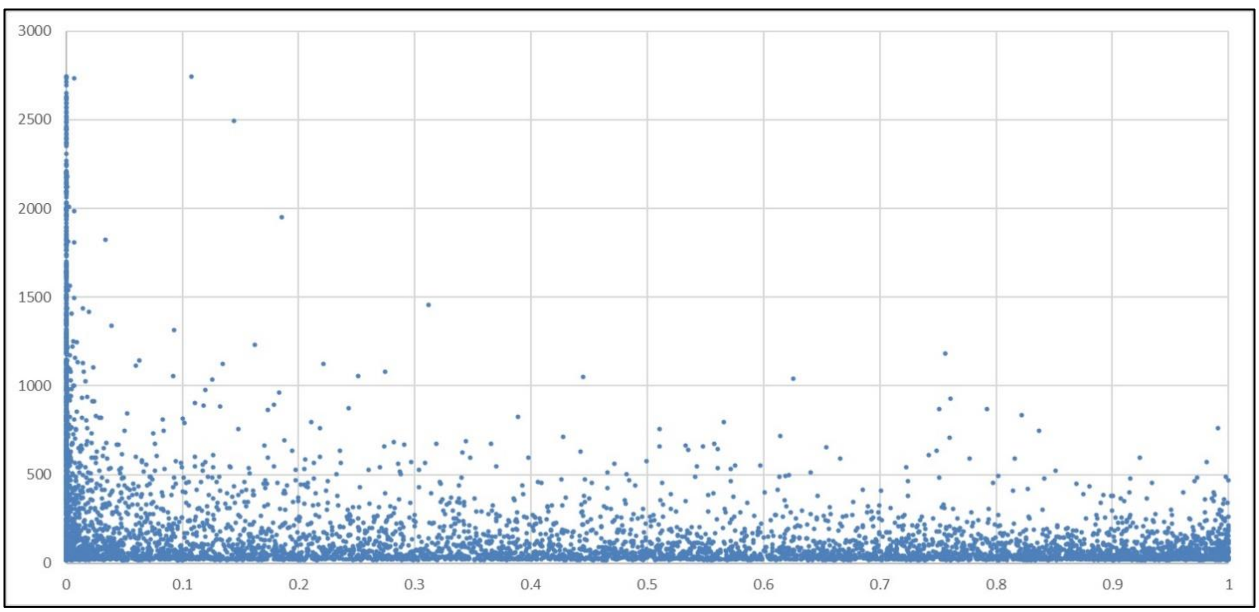
A closer inspection of a commonly performed billing codes of short (laparoscopic appendectomy), medium (total knee), and long (coronary artery bypass graft) case lengths shows that while some case lengths are normally distributed, it is because there’s so many physicians involved that some would end up normally distributed. Non-parametric testing will be used to compare case length populations: Kruskal-Wallis tests for means, and Levene tests for variances. The following charts show the results of those tests for providers using the same billing code with at least 20 cases in the data set.

Figure 4 Kruskal-Wallis Provider - Billing Code P Values by Case Count



Data from Vizient Procedural Analytics used with permission of Vizient, Inc. All rights reserved.

Figure 5 Levene Provider - Billing Code P Values by Case Count



Data from Vizient Procedural Analytics used with permission of Vizient, Inc. All rights reserved.

Like the normality testing, higher case counts led to lower p-values, indicating billing codes should be split by provider. The same testing by grouping physicians belonging to the same provider had similar results. Identifying meaningful differences between populations of case lengths and variances has two uses. First, to confirm that the same surgical procedure done at different hospitals may have different distributions, as well as different surgeons performing that procedure at the same hospital. Secondly, it shifts the focus away from thinking about the billing codes by the medical procedure and more on their statistical markers like mean and variance. Two different hospitals may schedule the same procedure, but if they use the same guidance based on that procedure it would ignore the differences in surgeon speed and each staff's checklist for case preparation. Focusing on the mean and variance of case lengths allows for completely different surgeries to be compared in similar terms regarding how they could be organized on a schedule.

Skewness is another concern for case lengths. There's a minimum amount of time needed to perform a specific surgery if everything lines up, but complications and unexpected issues occurring during the case could result in a much longer surgical time. This is borne out in the data, especially with physician and surgery types that were performed more frequently.

Table 6 Skewness of Physician Case Lengths by Times Performed

Case Count	Skew Under -1	Skew Between -1 and -.5	Skew Between -.5 and .5	Skew Between .5 and 1	Skew Above 1
5 And Under	1.1%	16.5%	48.4%	30.0%	4.0%
6-25	1.7%	5.3%	43.0%	24.8%	25.2%
26 And Over	.3%	.6%	19.6%	28.0%	51.4%

Data from Vizient Procedural Analytics used with permission of Vizient, Inc. All rights reserved. Skewness grouping from (Bulmer, 1979).

The issue with using skewness comparisons to relate cases is because long cases do happen, complications do occur, and it is more likely to happen at least once if a case is performed more often. Given that the idea of combining similar surgical procedures that happen infrequently is to make scheduling more accurate, grouping procedures based on similar number of outliers in the dataset means planning based on circumstances that operating room staff work to avoid. Skewness can be ignored for case comparisons.

Other Timing Intervals

So far, the focus has been on the length of time the patient is in the room. There is enough information in the data set to break the wheels in wheels out time to three different lengths: wheels in to cut, cut to close, and close to wheels out. But is there a reason to do so? Since operating room efficiency is based on the time the patient is in the room, and adding those three time intervals results together results in that case's wheels in wheels out time, there would have to be a compelling reason to change from using the total case time to three different time intervals. Either way, an overview of wheels in to cut and close to wheels out, the time of case setup and takedown surrounding cut-to-close, would be worth exploring.

Table 7 Distribution of Wheels In to Cut Case Lengths by Service Line, Ordered By Service Line Case Count

Service Line	Up to 15 Min	16-25 Min	26-35 Min	36-45 Min	46-60 Min	60+ Min
General Surgery	11.2%	32.5%	28.0%	14.6%	8.9%	4.8%
General Medicine	78.8%	13.0%	4.7%	1.9%	1.0%	0.5%
Orthopedics	10.9%	27.7%	31.4%	18.1%	9.1%	2.9%
Urology	17.5%	41.1%	22.1%	10.9%	6.2%	2.2%
Cardiology	21.7%	31.5%	19.7%	11.7%	8.7%	6.7%
Ophthalmology	64.1%	25.0%	7.6%	2.1%	0.8%	0.4%
Spinal Surgery	27.7%	7.5%	12.6%	14.8%	17.3%	20.0%
Gynecology	8.5%	33.6%	30.3%	16.2%	8.6%	2.8%
Vascular Surgery	8.2%	30.2%	26.6%	16.0%	11.2%	7.7%
Otolaryngology	30.8%	27.8%	20.6%	11.4%	6.4%	3.1%
Trauma	6.0%	25.2%	29.0%	19.3%	13.5%	7.1%
Neurosurgery	25.4%	10.7%	11.2%	10.6%	14.2%	27.9%
Plastic Surgery	5.2%	30.7%	32.7%	18.2%	10.0%	3.2%
Obstetrics	20.6%	46.8%	21.6%	7.1%	2.7%	1.3%
Cardiac Surgery	7.7%	9.0%	7.4%	8.7%	15.0%	52.2%
Thoracic Surgery	11.0%	12.1%	14.0%	15.7%	20.3%	26.9%

Pulmonary/ Critical Care	50.7%	29.6%	8.7%	4.2%	3.8%	2.9%
Dental/Oral Surgery	17.4%	42.9%	20.6%	8.9%	5.6%	4.7%
Transplant Services	0.4%	2.9%	11.1%	18.0%	23.3%	44.3%
Grand Total	26.4%	26.3%	20.9%	12.1%	8.1%	6.2%

Data from Vizient Procedural Analytics used with permission of Vizient, Inc. All rights reserved.

14 of the 19 service lines have the patient prepped for incision within 35 minutes over half the time. Quicker procedures, like colonoscopies in general medicine, cataract treatments in ophthalmology, and diagnosing lung issues in pulmonary care, form the bulk of the under-15-minute times, along with injections categorized in spinal and neurosurgery service lines. Procedures affecting the heart or brain tended to be longer due to the time required to induce full-body anesthesia. Otherwise, wheels in to cut took around 20 to 30 minutes. Being able to explain the wheels in to cut length of hundreds of thousands of cases over multiple services lines with “around 20 to 30 minutes” and having it not feel like an oversimplification is enough of a reason to stick to wheels in to wheels out. For completeness’s sake, here is a similar look at the close to wheels out time.

Table 8 Distribution of Close to Wheels Out Case Lengths by Service Line, Ordered By Service Line Case Count

Service Line	Up to 5 Min	6-10 Min	11-15 Min	16-30 Min	30+ Min
General Surgery	16.3%	32.5%	22.8%	21.7%	6.7%
General Medicine	64.8%	24.0%	6.7%	3.8%	0.7%
Orthopedics	34.0%	32.0%	15.2%	13.8%	4.9%
Urology	19.2%	38.8%	22.9%	15.7%	3.4%
Cardiology	19.5%	29.1%	21.8%	23.7%	5.9%
Ophthalmology	71.3%	16.7%	6.3%	4.6%	1.0%
Spinal Surgery	33.4%	17.9%	16.1%	22.7%	9.8%
Gynecology	15.6%	36.9%	24.5%	19.2%	3.9%
Vascular Surgery	18.5%	33.6%	20.5%	20.2%	7.2%

Otolaryngology	15.9%	25.6%	24.6%	27.5%	6.3%
Trauma	22.2%	27.0%	16.9%	21.9%	12.1%
Neurosurgery	28.7%	14.4%	14.1%	27.6%	15.2%
Plastic Surgery	17.2%	33.5%	20.2%	20.0%	9.1%
Obstetrics	18.9%	40.9%	21.9%	14.2%	4.1%
Cardiac Surgery	8.4%	17.5%	19.0%	34.1%	21.1%
Thoracic Surgery	8.4%	17.9%	22.2%	35.2%	16.4%
Pulmonary/ Critical Care	11.4%	30.4%	28.1%	24.7%	5.4%
Dental/Oral Surgery	9.8%	29.5%	27.3%	27.6%	5.9%
Transplant Services	2.0%	12.9%	23.0%	37.7%	24.4%
Grand Total	30.1%	28.6%	17.7%	17.6%	6.0%

Data from Vizient Procedural Analytics used with permission of Vizient, Inc. All rights reserved.

Three quarters of all patients are wheeled out of the operating room within 15 minutes of the close of the incision. The same exceptions for wheels in to cut hold here. There may be times when splitting a case would be useful, as the team at Massachusetts General did (Sokal, Craft, Chang, Sandberg, & Berger, 2006). But when looking at scheduling the entire case, using the time the patient is in the room is best, simplest, and most straightforward. Analysis based on one field is easier to explain than the three parts that add up to that one field.

An as-of-yet unexplored metric is turnaround time, which according to one survey showed a median turnaround time of 28.5 minutes. (Foster, 2012). Due to data issues around the actual procedural suite, the "ORDayCheck" flag was used to identify 15 providers that had a high number of cases that had valid turnaround times, resulting in a sample size of 264,858 cases. Looking at the kinds of cases that entered within two hours of the previous case leaving the room did show some useful trends.

Table 9 Average Turnaround Time based on Case Length and Prior Surgeon

Case Length Grouping	TAT When New Surgeon Enters Room	TAT When Surgeon Followed Self	All Cases
Over 120 Min WIWO	60.2 Min	41.7 Min	49.5 Min
40-120 Min WIWO	54.9 Min	31.0 Min	39.9 Min

Sub 40 Min WIWO	41.0 Min	18.4 Min	24.0 Min
All Cases	53.1 Min	28.3 Min	36.8 Min

Data from Vizient Procedural Analytics used with permission of Vizient, Inc. All rights reserved.

The fields, rows, and totals each contain insight into how turnaround time can be affected. First, if the surgeon is already in the room, the next case enters significantly quicker. Second, there is a relation between how long a case lasts and the amount of time it takes to set up for that case. Lastly, the average of all case turnaround times being closer to the turnaround times where a surgeon follows themselves shows the cases where a surgeon follows themselves outweighs the number of cases where another surgeon was previously in the room. There is a strong preference to have a surgeon stay in the same room for each successive case, reflecting a preference for block scheduling at the provider hospitals selected for turnaround analysis.

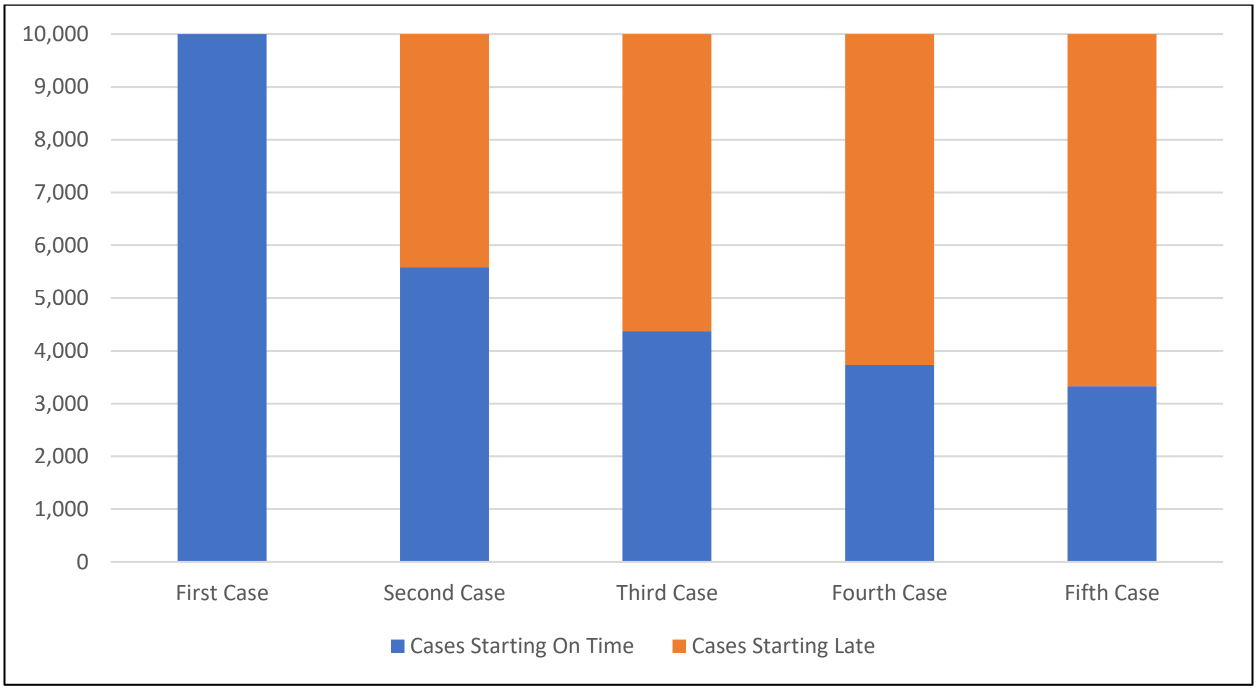
An Initial Simulation

Introducing a simplistic Monte Carlo simulation will help guide further analysis. Our initial simulation will be set up as follows:

1. Each simulated day in the operating room will have the same number of cases.
2. The same physician and surgery type will be used, in this case total knee replacements.
3. To reflect current accepted scheduling practices, the average of case lengths for the entire population will be used as the planned length of time. In other words, the second case will be scheduled to start at the mean case length, the third case will be scheduled to start at twice the mean case length, and so on.
4. There is no turnaround time between cases.
5. The next case only starts if the prior case has ended, and the scheduled start time has passed. All first cases start at exactly time 0.

One high-volume surgeon's pool of total knee replacements was repeatedly scheduled five times a day for 10,000 simulated days. The cases have a population mean of 119.4 minutes, so scheduled start times of 0, 120, 240, 360, and 480 were used for each case in the day. The following results summarize those simulations, starting with the timeliness of all five cases.

Figure 6 Simulated Cases Timeliness by Order in Day



Data from Vizient Procedural Analytics used with permission of Vizient, Inc. All rights reserved.

While the mean of the simulated cases is just under 120 minutes, it is clearly not the median as over half of all second cases started on time. Also, as the day progresses more cases start late. Looking at each simulated operating room day in a confusion matrix and observing how a case being on time or late affects the following case clarifies why.

Table 10 Confusion Matrix of Simulated Cases Starting on Time or Late

	Late Second Cases	On Time Second Cases
Late First Cases	0	0
On Time First Cases	4,418	5,582
	Late Third Cases	On Time Third Cases

Late Second Cases	3,153	1,265
On Time Second Cases	2,476	3,106
	Late Fourth Cases	On Time Fourth Cases
Late Third Cases	4,334	1,295
On Time Third Cases	1,932	2,439
	Late Fifth Cases	On Time Fifth Cases
Late Fourth Cases	4,998	1,268
On Time Fourth Cases	1,675	2,059

Data from Vizient Procedural Analytics used with permission of Vizient, Inc. All rights reserved.

Table 11 Cases Being On-Time or Late As Result of Prior Case

	Late Second Cases	On Time Second Cases
All Late First Cases	0	0
On Time First Cases	44.2%	55.8%
	Late Third Cases	On Time Third Cases
Late Second Cases	71.4%	28.6%
On Time Second Cases	44.4%	55.6%
	Late Fourth Cases	On Time Fourth Cases
Late Third Cases	77.0%	23.0%
On Time Third Cases	44.2%	55.8%
	Late Fifth Cases	On Time Fifth Cases
Late Fourth Cases	79.8%	20.2%
On Time Fourth Cases	44.9%	55.1%

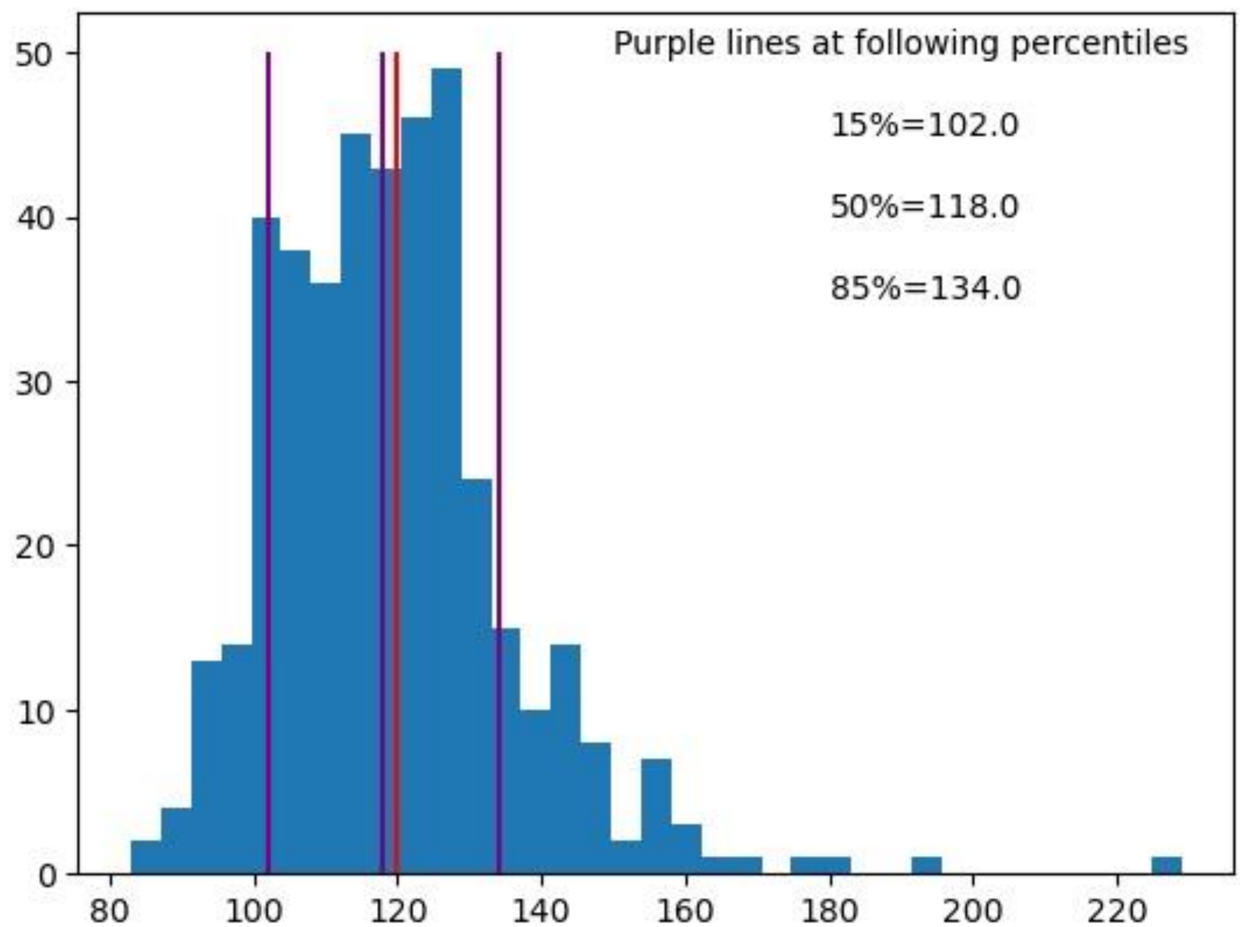
Data from Vizient Procedural Analytics used with permission of Vizient, Inc. All rights reserved.

Due to the restriction that a case does not start before the scheduled start time, any case that finishes early functions the same as if that case started the day. A case that starts on time has the same 55% chance of finishing early. The 45% of cases that finish late can make up for lost time and get back on schedule, but that happens less often as the day progresses.

Another issue emerges when looking at when the operating room day ends. With the final case of the day scheduled at 480 minutes, the only way a day could end on time or early is if the last case of the day was part of the 55% of cases that finished before 120 minutes, as well starting on time or only slightly late. In the simulation, only 3,000 days finished at or before the 600-minute mark. Of the other 7,000 days, 4,411 would not have finished within the allotted time due to the total wheels in-wheels out time exceeding 600 minutes. That leaves 2,589 simulated days that required extra time to finish, but due to the scheduling start requirement had total empty operating room time in the middle of the day that exceeded the minutes running over 600.

Counterbalancing that possibility is the simple fact that if all 5 patients had their surgeries scheduled at time 0, there would never be a gap between surgical cases. Instead, in each of those simulations there is one fictional patient waiting around for eight hours. Even in this simulation, using the average case length led to 22,986 late cases of the 40,000 non-first cases scheduled. Each hospital will adjust their own scheduling dial between “maximum operating room efficiency” and “crowded pre-op waiting rooms.” Looking at the distribution of case lengths used in the simulation will help form ideas around scheduling.

Figure 7 Histogram of Case Lengths Used in Simulation



Data from Vizient Procedural Analytics used with permission of Vizient, Inc. All rights reserved.

The mean isn't a bad estimate for the length of any single case. About 70% of all cases are within 16 minutes, the standard deviation of the population. The downside with using the average case length to plan a schedule is that it assumes a minute late is the same as a minute early and cedes control of the scheduling dial over to variance in case length. As shown in our simulation, it does a mediocre job of both avoiding empty operating room time between cases as well as avoiding delays. Additionally, the case lengths that fall outside of one standard deviation are dispersed differently. If a case was 102 minutes or less, it averaged 97 minutes. If it was 134 minutes or slower, it took on average 147.8

minutes. Weighing each circumstance depending on the stated business priority of the operating room, efficiency versus scheduling accuracy, will be a necessary addition to improve future simulations.

Improving the Simulation Model

The simple Monte Carlo simulation needs some refinements, starting with determining the starting case time for non-first cases. Of 1,193 physician-procedure combinations with at least 50 cases and a wheels-in wheels-out mean between 110 and 130, the simple simulation's variance was around the 20th percentile. Higher variances would mean more empty operating room time as well as more significant delays for cases.

Looking at the confusion matrix of cases showing when the next case was early or late based on if the preceding case was early or late showed that there were interactions between subsequent cases. While on-time starts consistently resulted in the next case also being on time around 55% of the time no matter what slot in the day (with empty OR time preceding it), the late start to late start pipeline kept growing each iteration. Why not simulate prior cases before scheduling the next case? Setting each case start time all at once based on average case lengths, while reflective of real-world practices, also resulted in the real-world issues of case delays increasing as the day progresses.

Table 12 Case Delays by Scheduled Hour of Day

Scheduled Hour of Day	On Time or Early	Late	Total Cases
7 AM	43.1%	56.9%	527,650
8 AM	40.5%	59.5%	368,686
9 AM	34.5%	65.5%	320,609
10 AM	33.9%	66.1%	322,949
11 AM	35.8%	64.2%	291,085
12 PM	33.8%	66.2%	282,355
1 PM	33.5%	66.5%	243,370
2 PM	37.4%	62.6%	168,627
3 PM	41.6%	58.4%	110,785
4 PM	47.3%	52.7%	61,024

5 PM	49.6%	50.4%	35,343
6 PM	53.0%	47.0%	21,246
Total	**37.9%**	**62.1%**	**2,753,729**

Data from Vizient Procedural Analytics used with permission of Vizient, Inc. All rights reserved.

These real-life delays are not necessarily because of earlier operating room cases. For example, pre-operation admissions and testing could delay the patient's arrival to the operating room (Obermair, 2016). To focus on case length variance relating to the existing schedule, the stipulation of the next case starting only when the prior case ends, and the scheduled start time is reached will be kept for all future simulations.

The fictional patients are prompt and ready. Ensuring that they don't wait longer than the guidelines set by fictional hospital management to keep fictional patient satisfaction high is another concern, and happier patients do lead to increased financial performance (Betts, Balan-Cohen, Shukla, & Kumar, 2016). Unfortunately, the fictional operating room still costs between $30 and $100 a minute to staff (CaseCtrl, 2021). But there is no need to wait for a fictional board meeting to determine if the priority should be optimal efficiency or on-time starts. Our simulations can incorporate a way to weigh the minutes of empty operating room time versus case delays at each non-first case starting time. There needs to be a penalty for both kinds of minutes to avoid scheduling every case at time 0 or ten hours apart. That penalty could be equal or weighted to one side or another to better control the scheduling dial. Additionally, this model of simulating multiple cases in a row allows for later case scheduling start times to be impacted by earlier cases.

There is a reason hospitals use the average case length for scheduling operating rooms instead of robust Monte Carlo simulations. A 2016 survey of 50 companies found that health care providers were among the lowest ranked industries in analytical competitiveness (Alles & Burshek, 2016). Going away from a readily available case length average would be difficult, but helpful, directed advice based on the simulations would be useful to nudge cases one way or another. Linear regression analysis will help summarize simulation results in a way that helps for easier understanding and application to operating room schedules. Additionally, these models can be checked by applying them to other facilities, creating schedules based on case order, means and variance. Then, the schedules those models can create can be run through a simulation to check the findings.

Creating and Simulating Operating Room Schedules

Two processes were created in Python to ensure all simulations were consistent. First, a method adds scheduled starting times to a series of valid physician-billing code combinations. The first case is scheduled to start at time 0. A randomly chosen sample of 100 case lengths, with replacement, is created from that physician-billing code combination. To find the next case's starting time, each possible time between the shortest and longest case lengths in the sample is considered. The impact of each possible start time is calculated by subtracting the start time by each simulated case result length, and multiplied by either the weight of a gap between cases for start times past the previous case end or the weight of the delay for a start time that occurs when the prior case is still occurring. The minimum result is the next scheduled start time. The 1000 simulations used to create each scheduled start time are carried forward, with any case finishing before the new scheduled start time changed to the new scheduled start time. This not only adjusts for the rule preventing cases starting before they are scheduled but helps proportionally weigh the earlier finishing cases.

The second process takes those physician-billing code combinations and start times and simulates 1000 different results of that operating room day. Each individual simulation has the gap or delay related to each non-first case recorded, and each is separately totaled for that simulation as well as the total case time and when each case leaves the operating room.

The Cutting Room Floor

Using a linear model to determine the length of time something will take is reasonable. But, using it to predict the length of a case is not an option in this data set due to the limited fields present. There's clearly differences in case length averages and variance between some physicians performing the same billing code. But that's all the relevant information included in the data set. Linear models need more information to go on to be effective. More importantly, there is no need to predict a case length when there's so many real case lengths to use. Monte Carlo simulations are a better fit for the data set. Linear regression analysis will be useful to see trends in the data from simulation results.

But randomly choosing from all cases would lead to some simulations choosing the same case length every time if a physician performed it once. To avoid including a constant into deterministic simulations, only combinations of a physician and billing code with at least five cases were included. Billing codes relating to services lines with small total case counts like oncology or codes relating to anesthesia were also removed.

Case length skewness has been discussed, but there are still some unrealistic case lengths that could be from inaccurately recording the timestamps or applying the wrong billing code to the relevant procedure. Based on the skewness review and looking at the histogram of case lengths used in the simple simulation, cases that were shorter than two standard deviations below the mean or longer than three standard deviations above the mean for all examples of that billing code were removed.

Table 13 Final Data Set Overview

	Total Cases	Billing Codes	Provider Count	Physician Count
General Medicine	337,414	158	105	1,359
General Surgery	330,198	926	115	3,970
Orthopedics	255,099	602	113	1,975
Ophthalmology	153,969	175	81	854
Cardiology	148,197	171	84	1,143
Urology	143,290	204	104	1,122

Spinal Surgery	114,253	226	91	887
Gynecology	98,145	124	103	1,568
Otolaryngology	94,672	226	93	986
Vascular Surgery	71,432	299	107	1,209
Plastic Surgery	65,906	52	101	703
Neurosurgery	64,580	272	89	952
Trauma	59,466	264	101	1,141
Obstetrics	41,168	25	91	1,167
Cardiac Surgery	30,073	99	68	478
Pulmonary/ Critical Care	16,288	43	60	334
Thoracic Surgery	15,342	131	65	427
Transplant Services	12,435	16	23	232
Dental/Oral Surgery	11,403	38	59	260
Totals	2,063,330	4,051	120	12,142

Data from Vizient Procedural Analytics used with permission of Vizient, Inc. *All rights reserved.*

The 40,409 too-long cases and 1,964 too-short cases were removed after their relation to other cases in the data set was established, so the 'ORCheckFlag' for establishing turnaround time and the order of cases a surgeon performed in a day may refer to cases not included in the final data set. This benefits the purpose of both fields, as the ORCheckFlag is a guard against bad data and the surgeon case order can be used to verify that all cases performed are accounted for. The research and analysis in the next chapter is based on cases in this finalized data set.

To focus on the uncertainty and variance of case lengths, turnaround times will not be included in simulations. Turnaround times were calculated from the wheels-out of the prior case to the wheels-in of the next case, not when the room was ready for the next case. Additionally, Table 9 showed several variables that affected the turnaround time between cases. While simulations should strive for realism, including another variable would complicate findings with little benefit beyond an attempt at authenticity.

Chapter 4 –Results

Three kinds of operating room simulations were run. The first is a single physician-billing code combination. With the groundwork of An Initial Simulation already laid, the results of simulations using that same physician performing total knee replacements using different weighing penalties is discussed before showing similar findings for additional physician-billing code combination simulations. Then, multiple different billing codes are looked at in the same light before looking at how changing the ordering of cases can change how the simulated operating day proceeded.

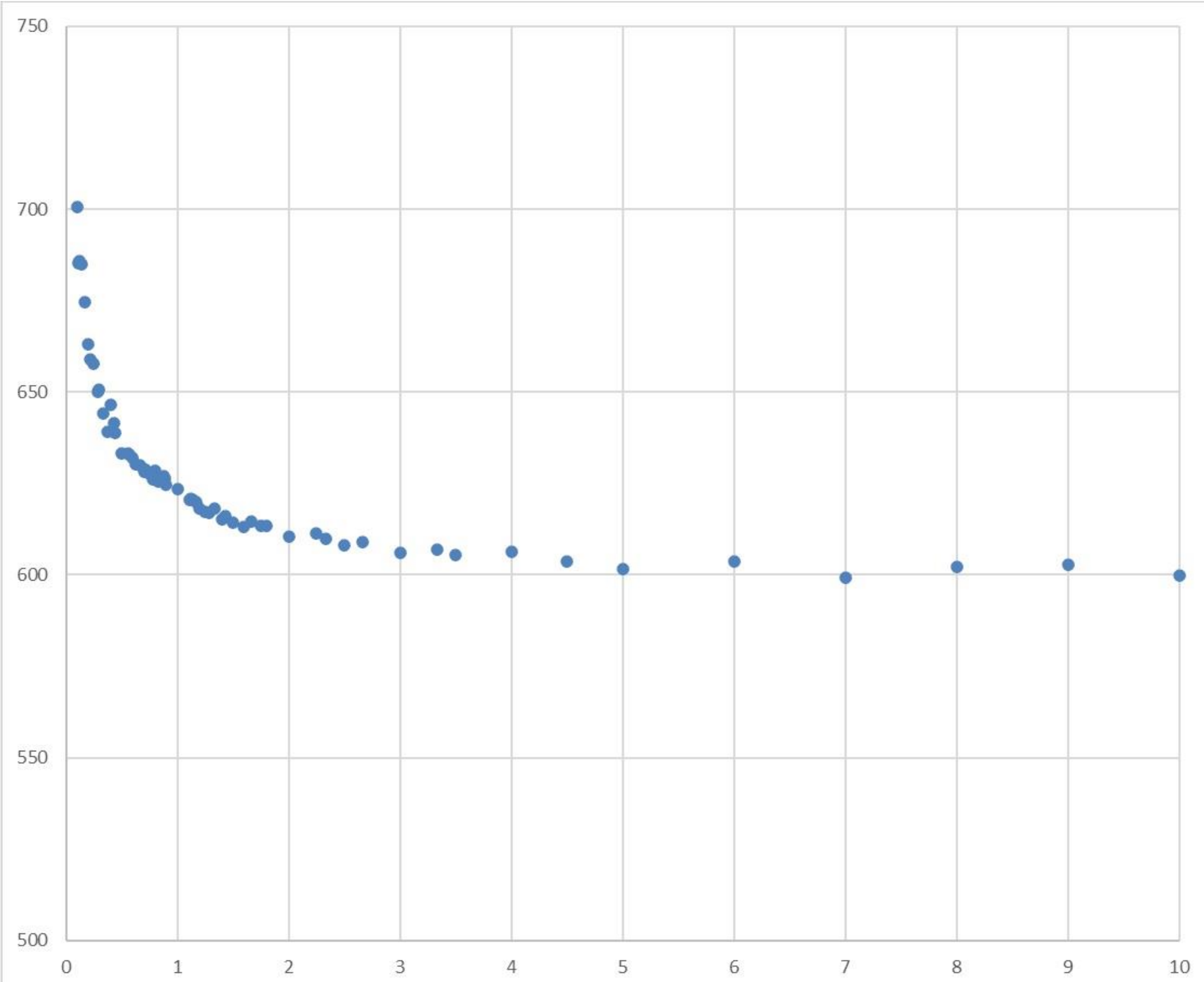
Comparing scheduling techniques requires creating metrics to summarize case delays and empty operating room time. Case delays will be compared by the total length of those delays. Empty operating room time means completing the same cases in a longer amount of time. Rather than use operating room utilization, empty operating room time will be monitored by when the last case is completed.

A Complicated Simulation

Using sixty-three different combinations of weights for penalizing empty operating room time and delayed cases, sixty-three different schedules for the same total knee replacement performed by the same surgeon as the initial simple simulation were created. Weights between 1 and 10 were used for both case gaps and case delays, except when a particular combination would reduce to the same ratio. In the following graphs, the ratio is shown as the case gap weight divided by the case delay weight, meaning .1 is the strongest preference for avoiding delays and 10 is the strongest preference for avoiding empty operating room time. Each weighted schedule was simulated 1,000 times, and the same metrics around case delays and final case completion were derived. For reference, the simple simulation had an average time of 616.6 minutes to complete all five cases with an average total delay of 46.9 minutes.

The simulated case delays show two relationships to the priorities used in scheduling. The average finishing time decreased as the weight of minutes for penalizing empty operating room time increased (End of case time = -19.52 * ln(early penalty weight / late penalty weight) + 627.68, R^2 = .8787).

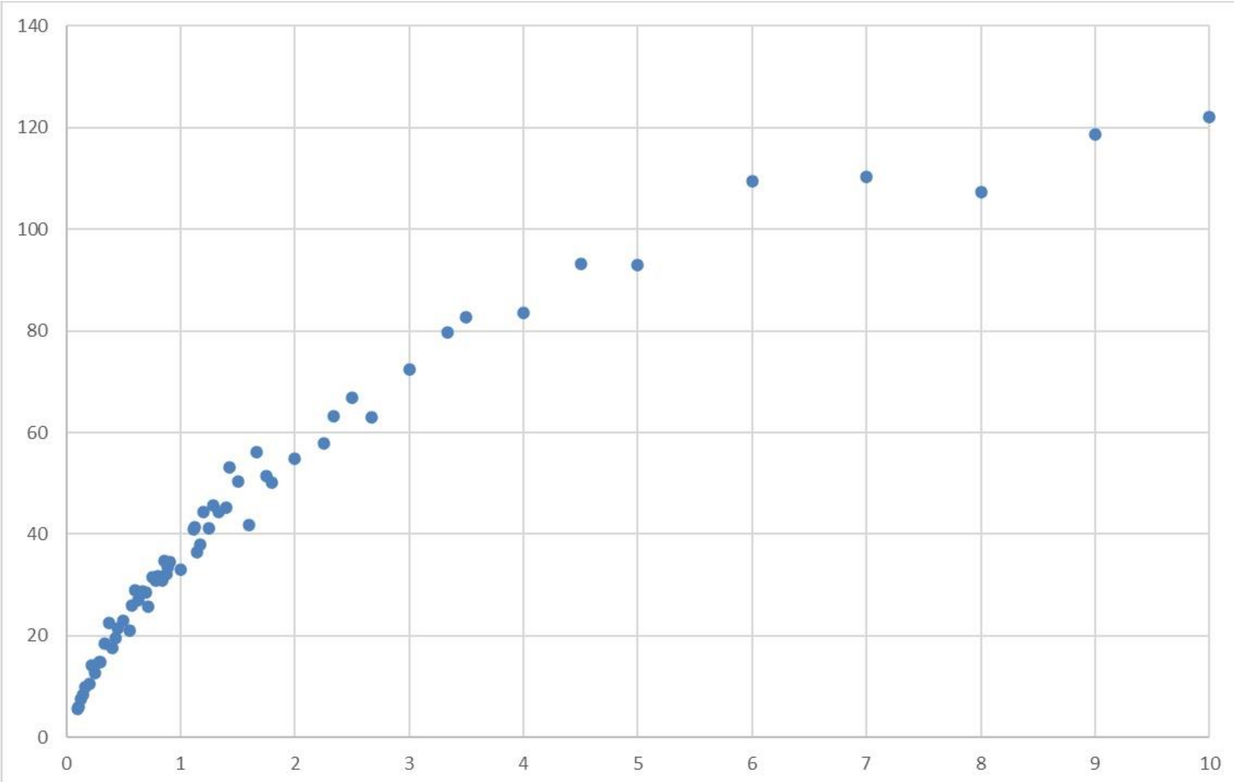
Figure 8 Average End of Simulated Knee Surgeries by Weighted Scheduling



Data from Vizient Procedural Analytics used with permission of Vizient, Inc. All rights reserved.

The increasing ratio has the opposite effect on the case delays (Total avg delay = 25.07 * ln(early penalty weight / late penalty weight) + 43.893, R^2 = .907).

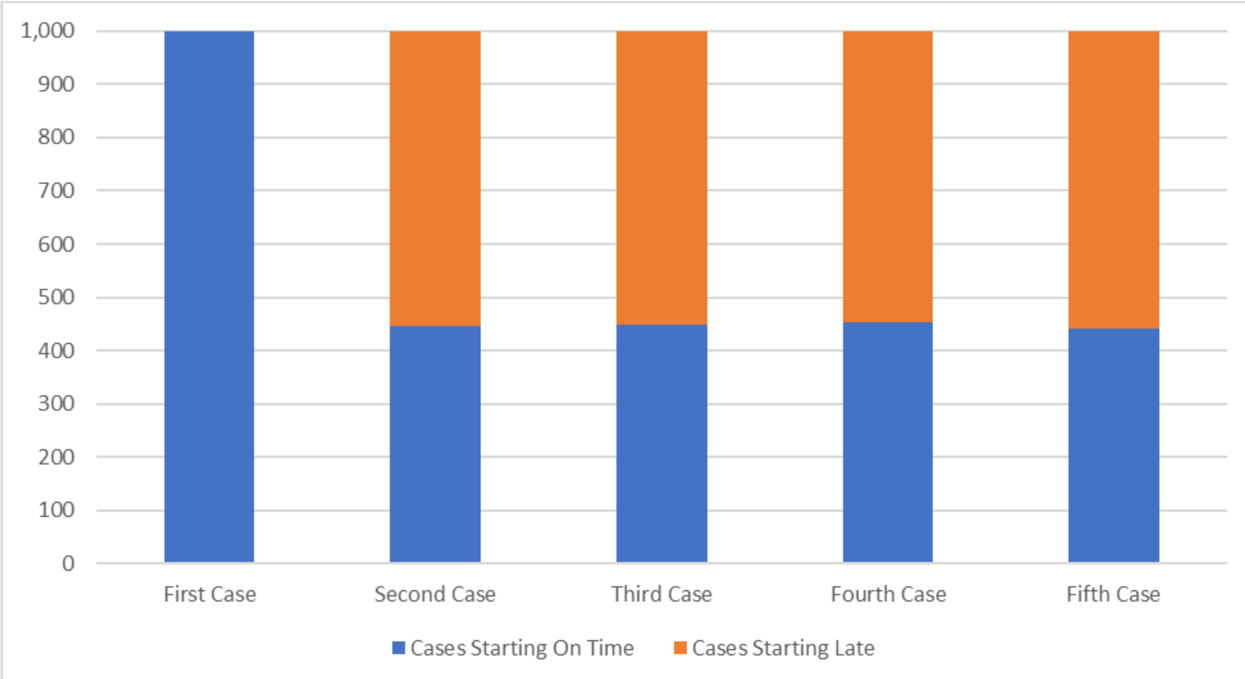
Figure 9 Average Total Delays of Simulated Knee Surgeries by Weighted Scheduling



Data from Vizient Procedural Analytics used with permission of Vizient, Inc. All rights reserved.

The percentage of delays were balanced across the non-first cases for all ratios. None of the 63 ratios used had a standard deviation above 30 for the total amount of delays in all simulations across the four time slots. Comparatively, the simulations in Figure 6 Simulated Cases Timeliness by Order in Day that were scheduled by average case length had a standard deviation of 85.25 after adjusting for the difference in number of simulations.

Figure 10 Weighted Simulated Total Knee Replacement Case Timeliness with Ratio 1.6 By Order In Day



Data from Vizient Procedural Analytics used with permission of Vizient, Inc. All rights reserved.

Scheduling these total knee replacements by case length average led to total delays of 49.35 minutes. All weighted ratios of 1.4 and below had less delays on average, as well as 1.6. Scheduling by case length average led to an average final wheels-out at 613.19. In addition to 1.6, each ratio at 2 or above also improved upon the baseline. The ratios 1.429, 1.5, 1.667, 1.75, and 1.8 performed worse in both metrics than the baseline. The quickest average final case was 599.32 at ratio 7/1, which had an average total delay of 110.37 minutes. Multiple ratios had average delays under one minute, with the final case finishing after 650.07 minutes on average in those simulations.

The scheduled start times were set by the weighted penalties for empty operating room time and late cases each simulation. The created schedules show differences in the time allotted for the first case of the day compared to the following cases.

Table 14 Time Allotted for Total Knee Replacement Cases in Schedules by Ratio

Ratio Used	Time for 1st Case	Time For 2nd Case	Time for 3rd Case	Time for 4th Case

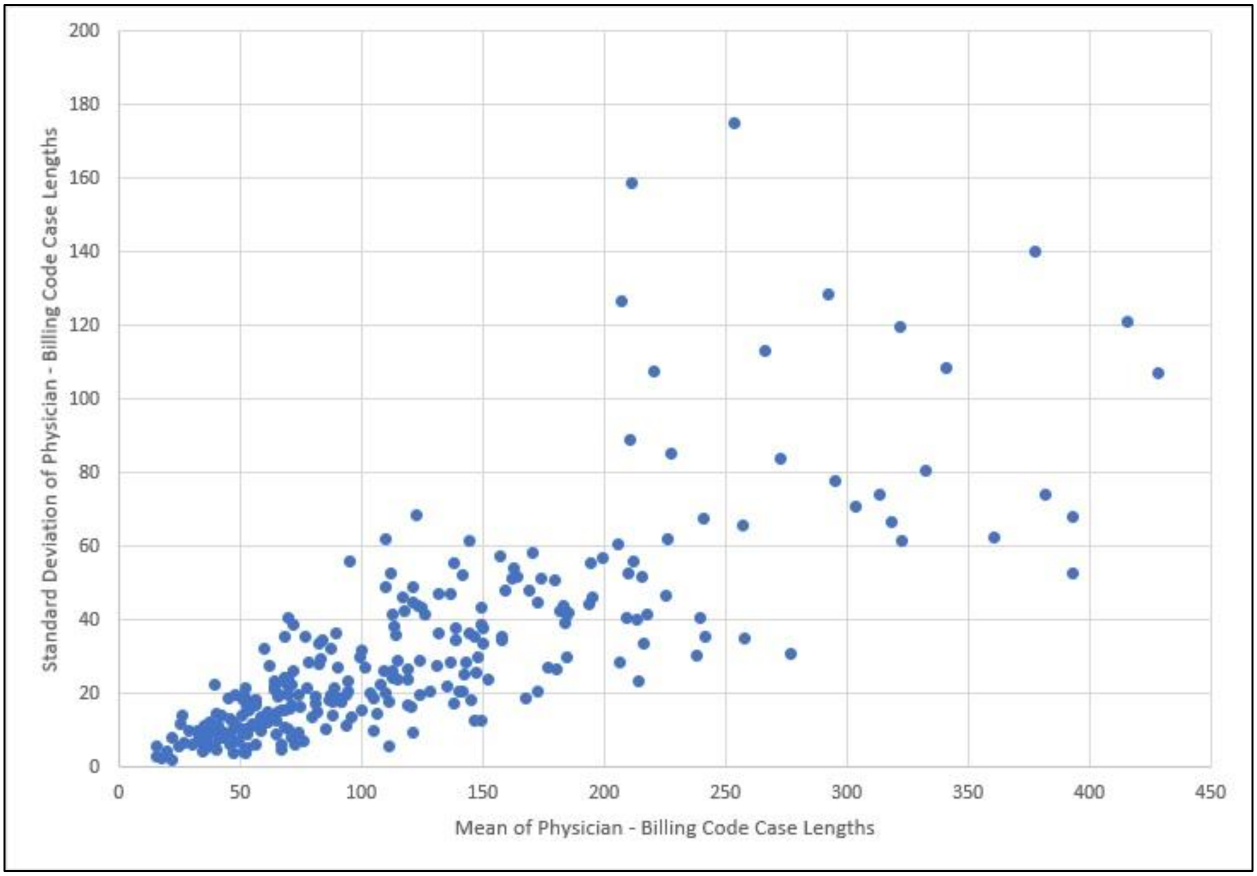
.5	123	126	129	130
1.5	115	120	122	123
2.5	110	117	120	122
3.5	105	118	118	120
4.5	105	115	117	119

Data from Vizient Procedural Analytics used with permission of Vizient, Inc. All rights reserved.

More Weighted Simulations with One Physician and Billing Code

This simulation was run on 276 additional randomly chosen physician-billing code combinations.

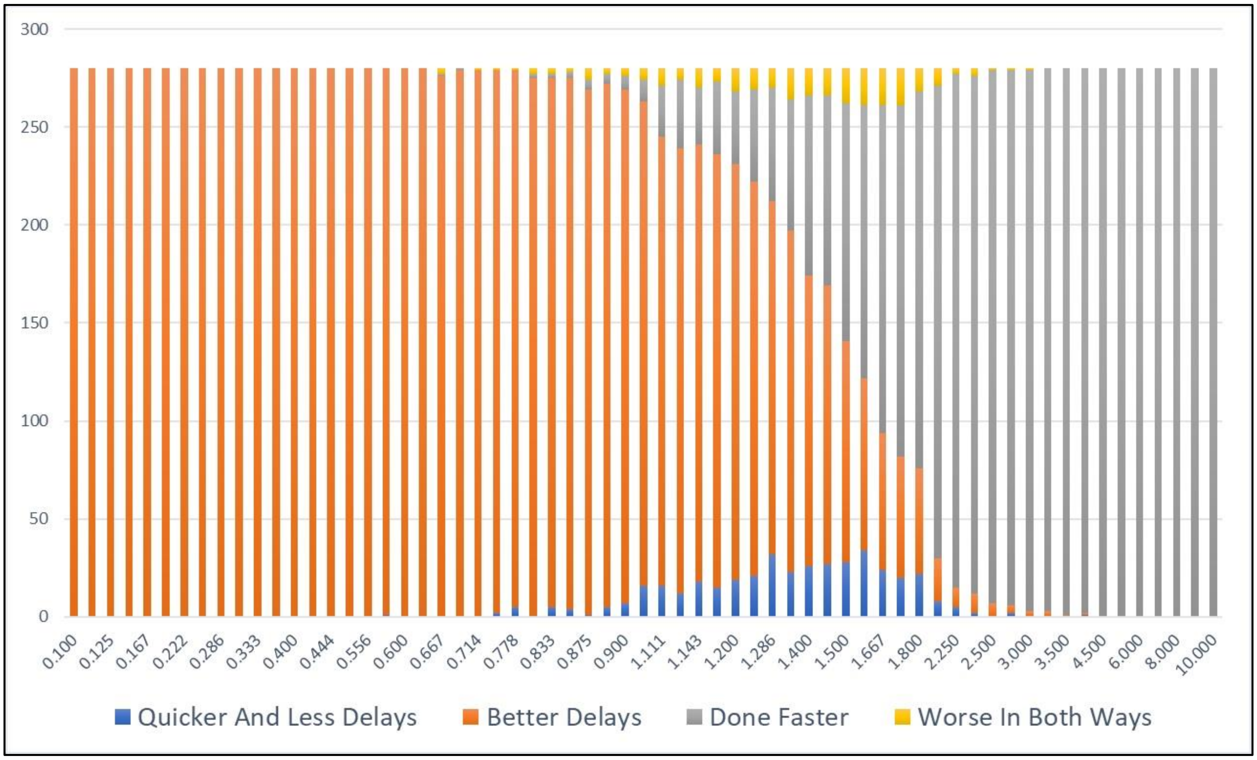
Figure 11 Mean and Variance for Additional Weighted Simulation Physician-Billing Codes



Data from Vizient Procedural Analytics used with permission of Vizient, Inc. All rights reserved.

The same schedules based on average case lengths were created for each physician-billing code combination to form the baseline, and the same 63 ratios for weighing schedules were used. The resulting performance in the two metrics compared to baseline for each simulation is shown below.

Figure 12 All Single Billing Code Simulation Results Compared to Baseline

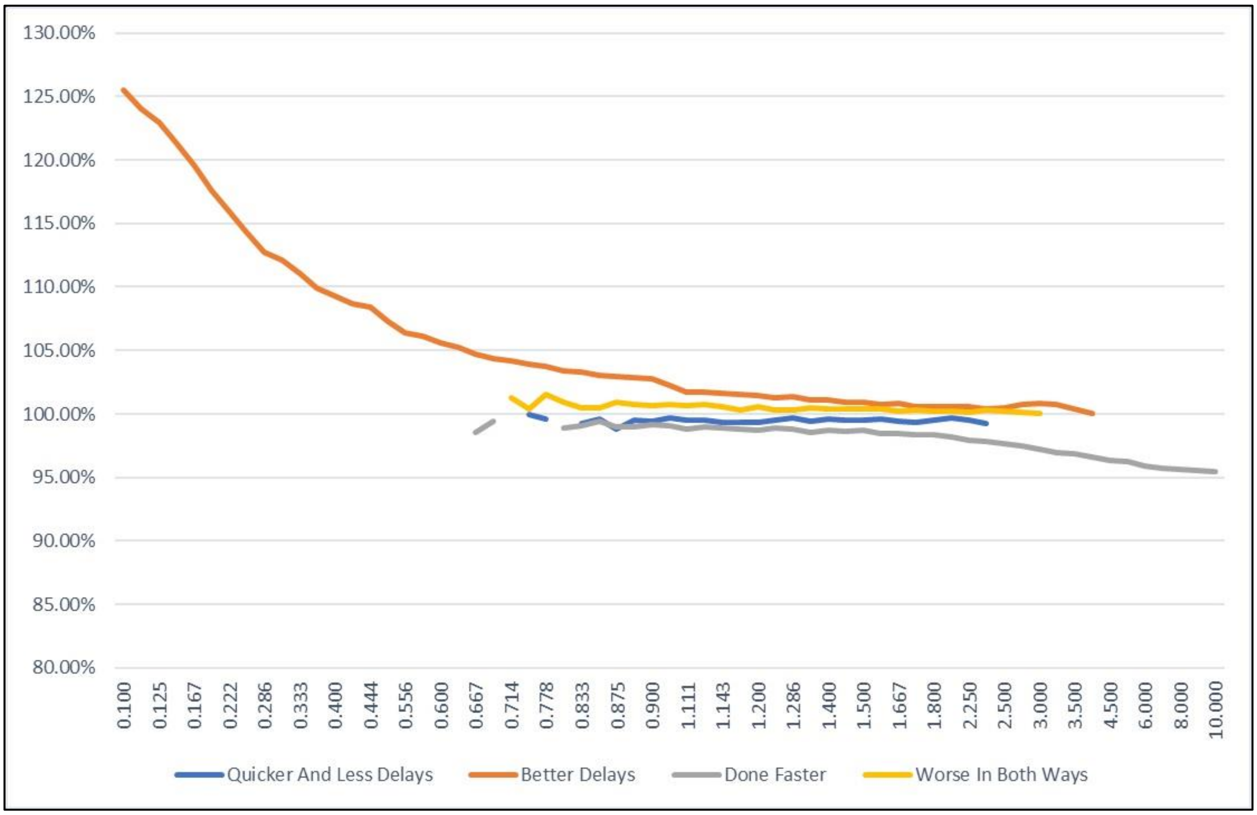


Data from Vizient Procedural Analytics used with permission of Vizient, Inc. All rights reserved.

Of the 17,640 simulations, 401 (2.3%) were better than the baseline in total delays and when the final case finished, and 248 (1.4%) performed worse in both. 161 physician-billing code combinations had at least one better performing simulation, compared to 113 with at least one worse-performing simulation. 55 had a simulation with both types of results. All ratios where the simulation either outperformed the baseline in both metrics or were worse in both metrics were between .571 and 3.

Dividing each simulation's final case end by the matching baseline's final case to create a percentage allows for comparisons across case lengths to help understand what happens in that span of simulated weighing.

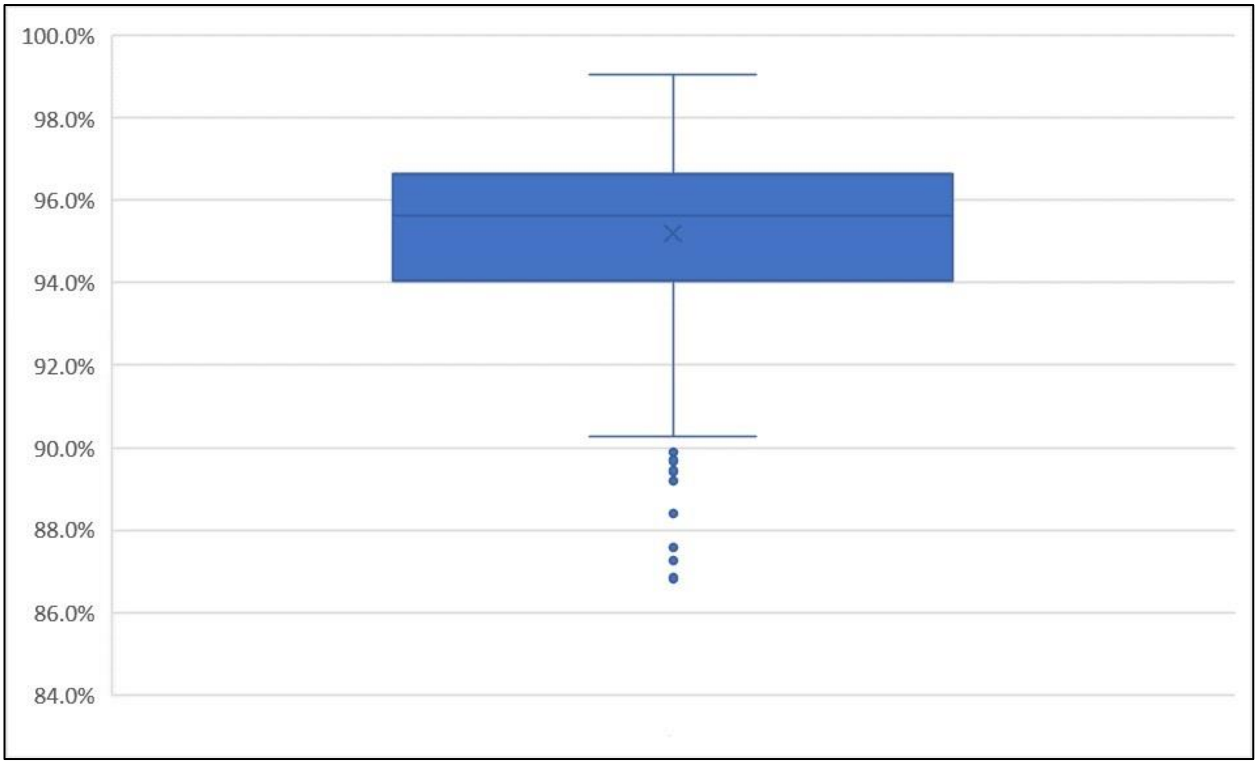
Figure 13 Average of Length of Operating Day by Simulation over Baseline for Single Case Simulations



Data from Vizient Procedural Analytics used with permission of Vizient, Inc. All rights reserved.

Each physician-billing code pairing did have at least one simulation that outperformed the baseline, generally when the early weight was at 10 (99 pairings), 9 (83 pairings), 8 (46 pairings) or 7 (33 pairings).

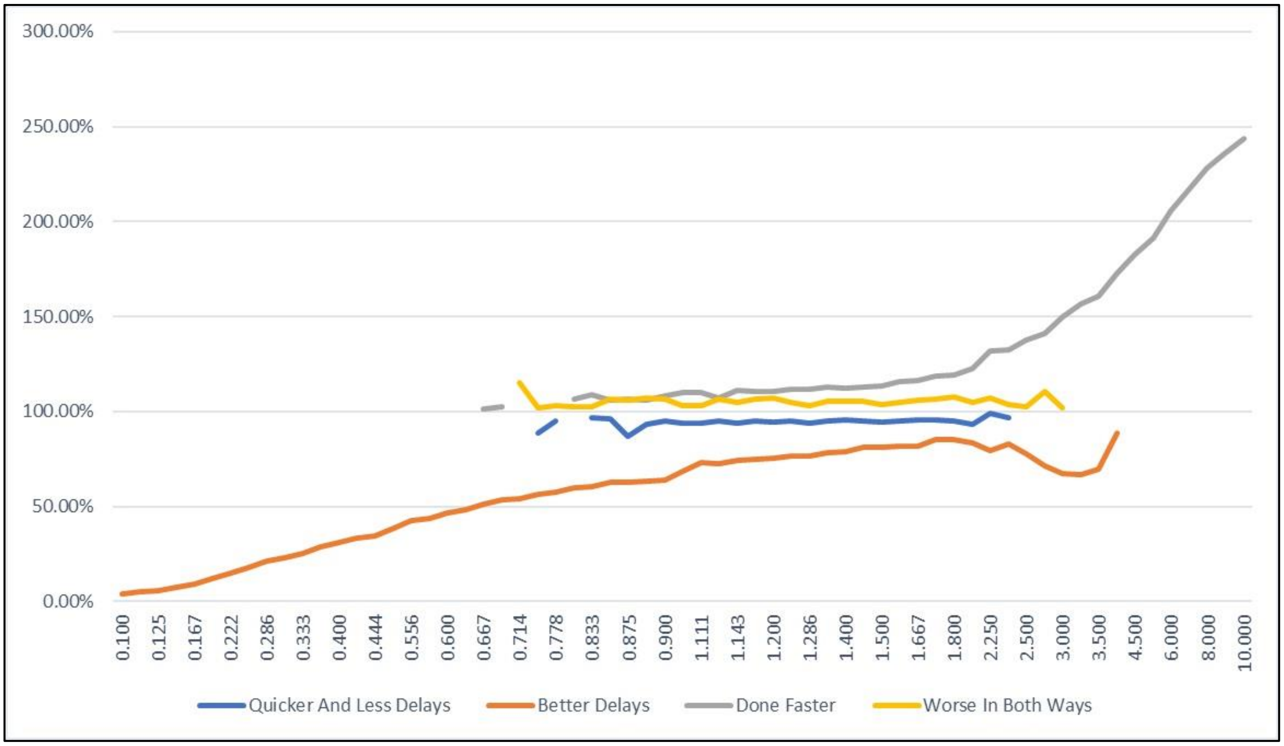
Figure 14 Best Performing End Times Compared to Baseline, All Single Billing Code Simulations



Data from Vizient Procedural Analytics used with permission of Vizient, Inc. All rights reserved.

The ratios that resulted in the last case of the day completing quicker also resulted in the most significant total case delays.

Figure 15 Average of Total Case Delays over Baseline, All Single Billing Code Simulations



Data from Vizient Procedural Analytics used with permission of Vizient, Inc. All rights reserved.

With different means for each physician-billing pairing in the simulation, using the gaps between scheduled case times as in Table 14 Time Allotted for Total Knee Replacement Cases in Schedules by Ratio wouldn't be a useful starting point to compare across different cases. A sample of simulation schedule gaps have been converted to the percentile of how it would compare to all relevant case times.

Table 15 Scheduled Case Lengths as Percentile of Cases in Billing Code Simulations

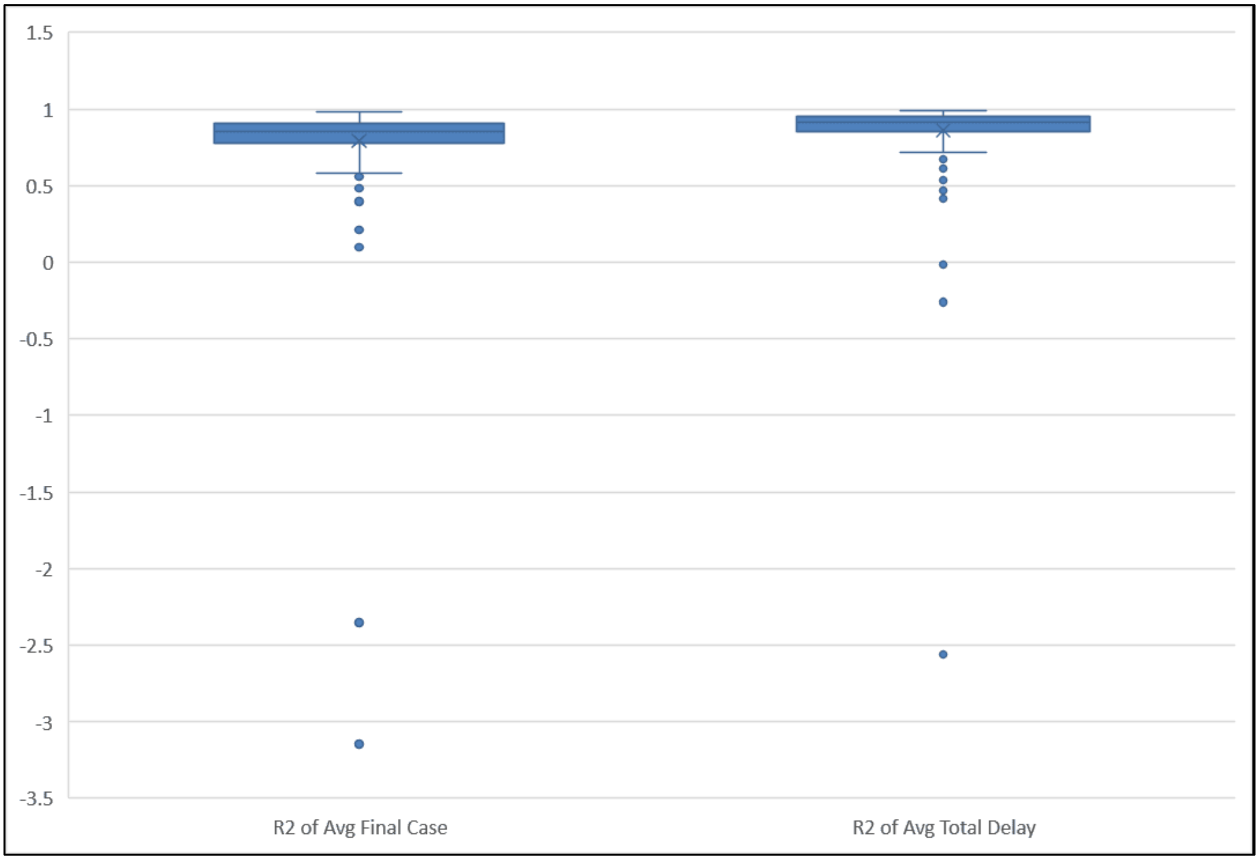
Physician – Billing Pairing	Weighing Ratio	1st Case Gap as Case Percentile	2nd Case Gap as Case Percentile	3rd Case Gap as Case Percentile	4th Case Gap as Case Percentile
A	5	9.8	56.4	32.0	59.4
B	5	12.1	45.5	57.6	57.6
C	5	33.3	48.1	48.1	63.0
D	5	12.1	42.4	63.6	57.6
A	3	23.7	53.4	32.0	56.4

B	3	12.1	45.5	53.0	65.2
C	3	24.1	59.3	59.3	70.4
D	3	18.2	47.0	57.6	57.6
A	1.667	42.9	53.4	53.4	53.4
B	1.667	36.4	48.5	57.6	81.8
C	1.667	33.3	66.7	66.7	77.8
D	1.667	47.0	66.7	57.6	57.6
A	1	42.9	50.4	73.3	63.5
B	1	60.6	65.2	71.2	65.2
C	1	42.6	77.8	85.2	77.8
D	1	51.5	47.0	66.7	63.6
A	0.6	63.5	80.5	81.2	81.2
B	0.6	65.2	75.8	78.8	75.8
C	0.6	66.7	74.1	85.2	77.8
D	0.6	75.8	69.7	69.7	66.7
A	0.333	70.3	76.7	80.5	84.2
B	0.333	71.2	78.8	90.9	86.4
C	0.333	77.8	85.2	81.5	85.2
D	0.333	75.8	84.8	75.8	84.8
A	0.2	86.1	87.2	83.1	86.1
B	0.2	75.8	78.8	86.4	90.9
C	0.2	88.9	85.2	92.6	85.2
D	0.2	92.4	81.8	87.9	81.8

Projecting Results by Linear Regression

The weighted total knee replacement's scheduling ratio strongly correlating with the resulting time of the final case ending and total delays (Figure 8 and Figure 9). To further investigate this relationship, each physician and billing code pairing had their simulation results split into a training set of 47 ratio and a testing set of 16 ratios. The resulting linear regressions showed a strong predictive ability for the resulting metric using only the ratio, with a few glaring exceptions.

Figure 16 Coefficient of Determinations from Linear Regression Modeling of Metrics Based on Ratio



Data from Vizient Procedural Analytics used with permission of Vizient, Inc. All rights reserved.

Projecting Linear Regressions by Linear Regression

Each of the resulting linear regression equations estimating the average total delays and average final case end time use a single variable, a ratio weighing case delays against empty operating room time. After recreating the model using all simulations for each physician-billing code combination, the coefficient of the ratio and the intercept that formed the prior 275 linear equations were each set as the dependent variable in two more linear equations. Each combination of mean, variance, skewness, and kurtosis were created using a model of 207 linear equations and tested with 70 linear equations held in reserve. The best combination of independent variables for each part of the linear equation for each metric are below, as well as the result of using the mean and variance if it was not the best variable combination.

Table 16 Results of Linear Regressions on Prior Simulation Linear Regressions

Linear Regression	**Metric**	**Stats**	**R2**
Intercept	Avg Final Case	Mean, variance	.9992
Ratio Coefficient	Avg Final Case	Mean, variance, kurtosis, skewness	.9521
Ratio Coefficient	Avg Total Delay	Mean, variance, kurtosis	.9246
Intercept	Avg Total Delay	Mean, variance, skewness	.9205
Ratio Coefficient	Avg Total Delay	Mean, variance	.9066
Ratio Coefficient	Avg Final Case	Mean, variance	.9055
Intercept	Avg Total Delay	Mean, variance	.8997

Data from Vizient Procedural Analytics used with permission of Vizient, Inc. All rights reserved.

Simulations Using Different Billing Codes

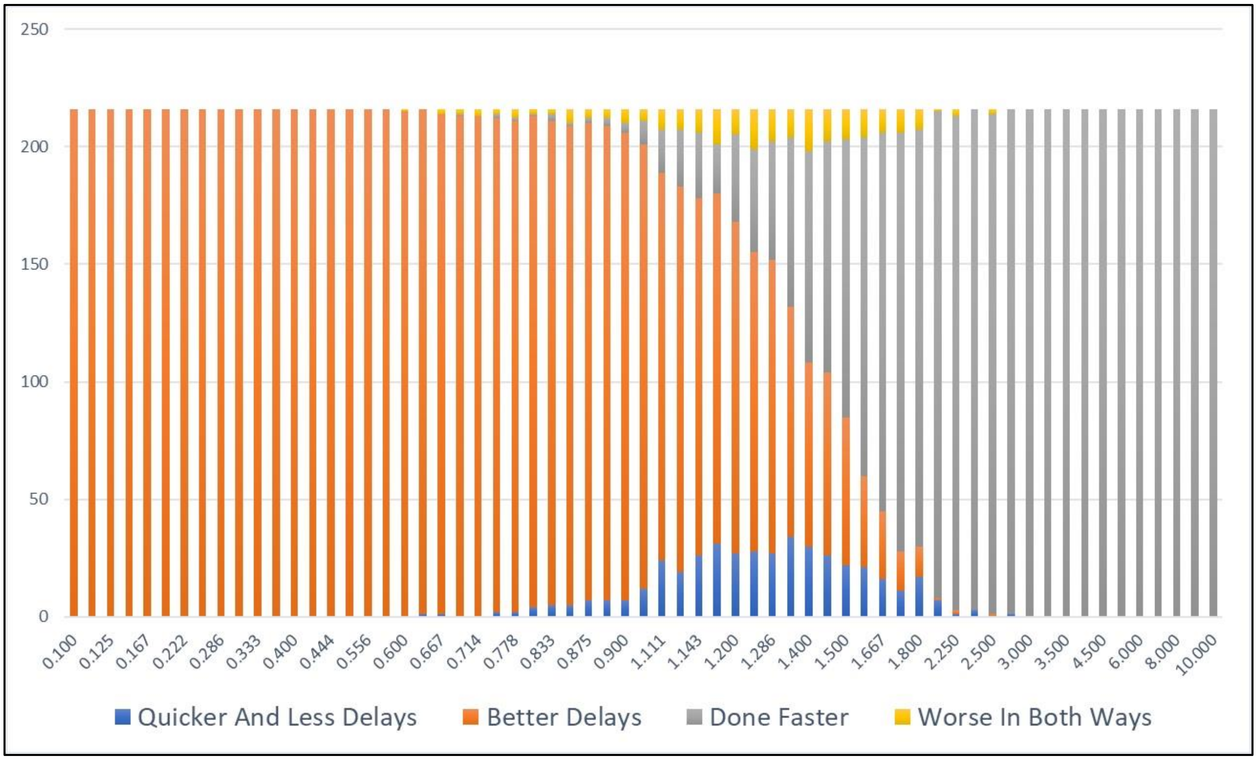
The same techniques were used on 216 schedules with multiple billing codes. Each schedule was created by randomly selecting one physician, then five different billing codes with resampling. The same order of each case was used in scheduling the start times based on all 63 weighted ratios and the baseline schedule using each case time's average length.

Table 17 Multiple Billing Code Simulation Overview

Number of Different Billing Codes	2 Codes	3 Codes	4 Codes	5 Codes	Total Simulations
Schedule Count	46	66	59	45	216

Data from Vizient Procedural Analytics used with permission of Vizient, Inc. All rights reserved.

Figure 17 All Multiple Billing Code Case Simulation Results Compared to Baseline



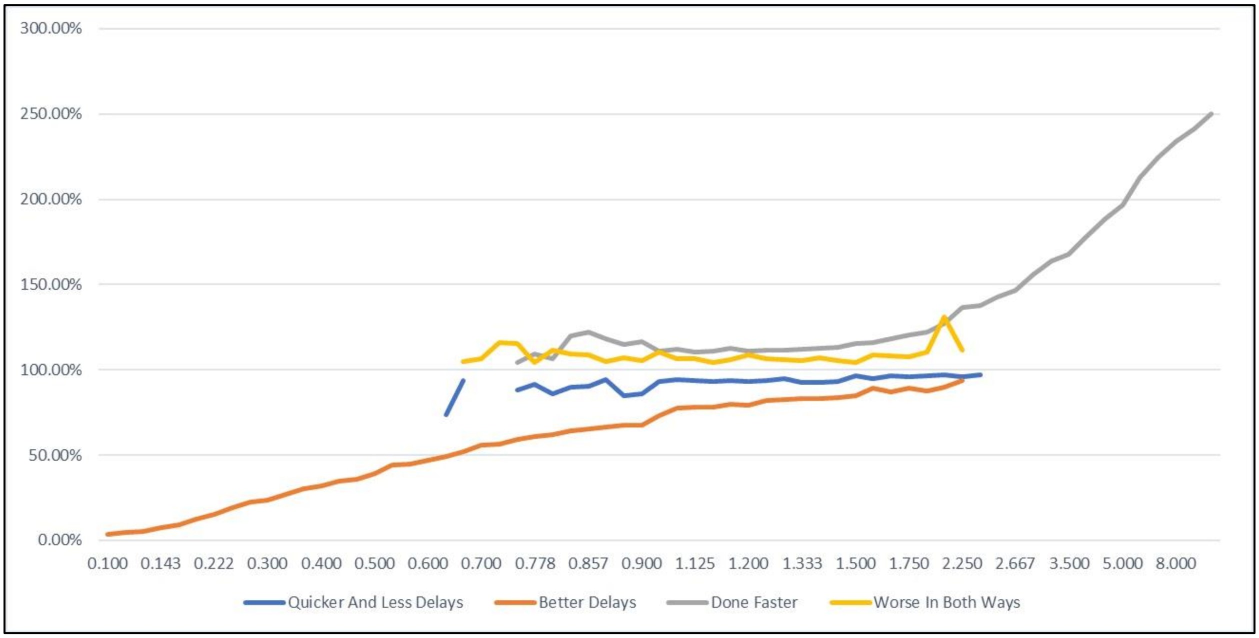
Data from Vizient Procedural Analytics used with permission of Vizient, Inc. All rights reserved.

Of the 13,608 simulations, 424 were better than the baseline in total delays and when the final case finished, and 232 performed worse in both. 125 physician-billing code combinations had at least one better performing simulation, compared to 98 with at least one worse-performing simulation. 149

had a simulation with both types of results. All ratios where the simulation either outperformed the baseline in both metrics or were worse in both metrics were between .6 and 2.667.

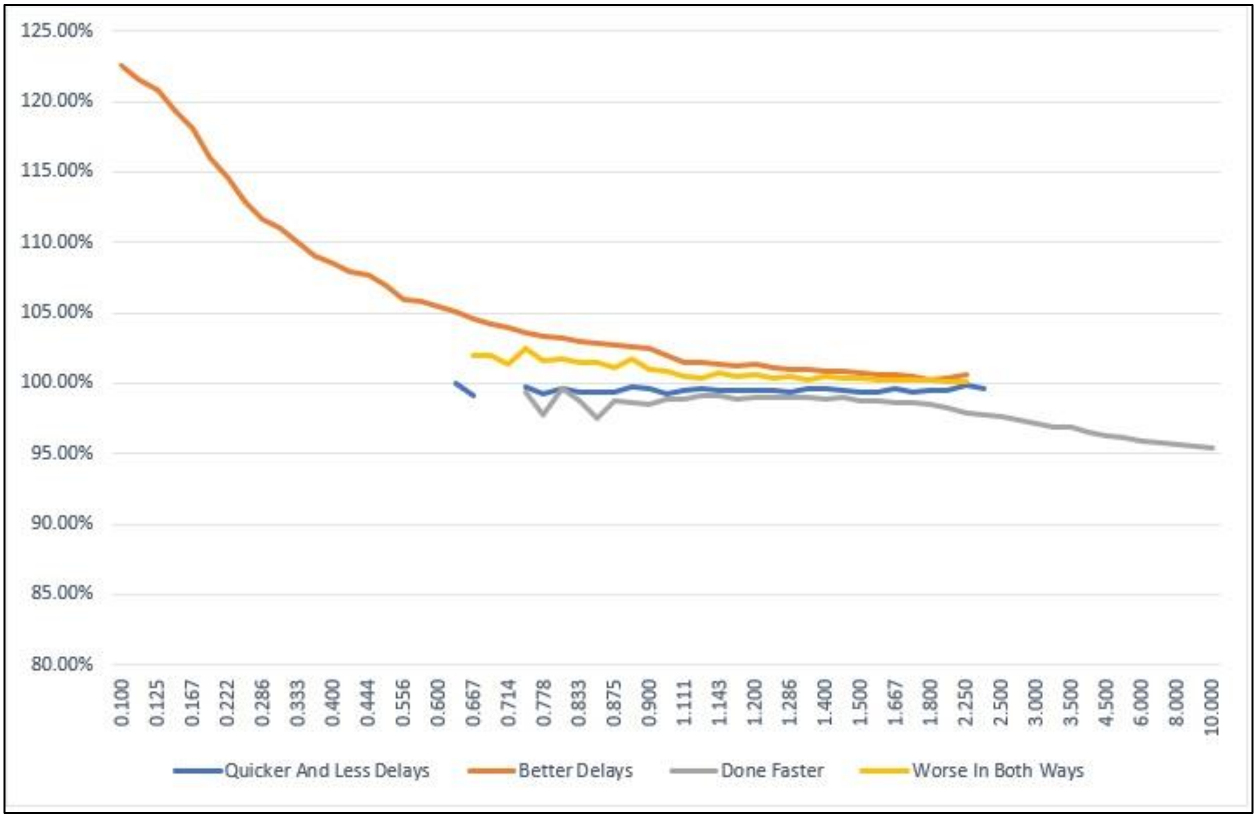
The inverse relation between case delays and last case ending time shown in the single billing code simulations is also found in the multiple billing code simulations.

Figure 18 Ratio of Total Case Delays over Baseline, All Multiple Billing Code Simulations by Ratio



Data from Vizient Procedural Analytics used with permission of Vizient, Inc. All rights reserved.

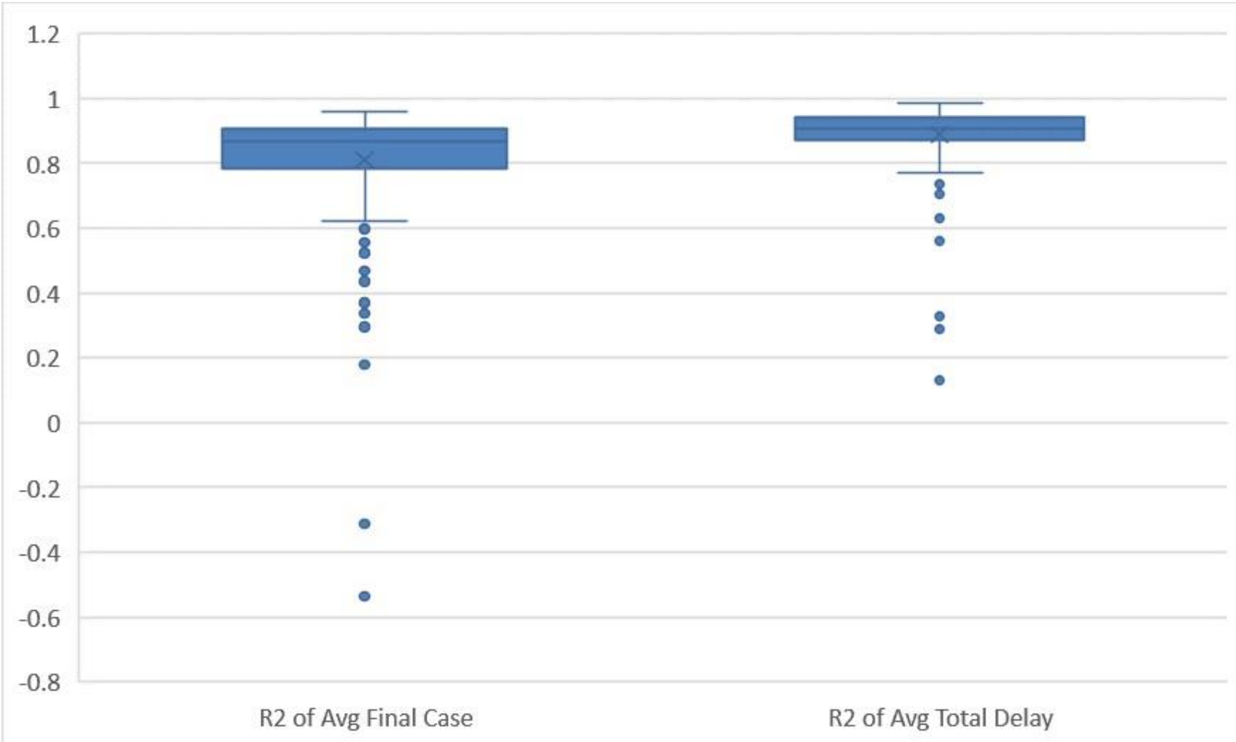
Figure 19 Ratio of Final Case End Times Compared to Baseline, All Multiple Billing Code Simulations by Ratio



Data from Vizient Procedural Analytics used with permission of Vizient, Inc. All rights reserved.

When using the same linear regression modeling approach as the single billing code simulations, the coefficient of the ratio in linear regressions for total delays was between 6.056 and 285.347, and for final case predictions was between -162.871 and -5.719.

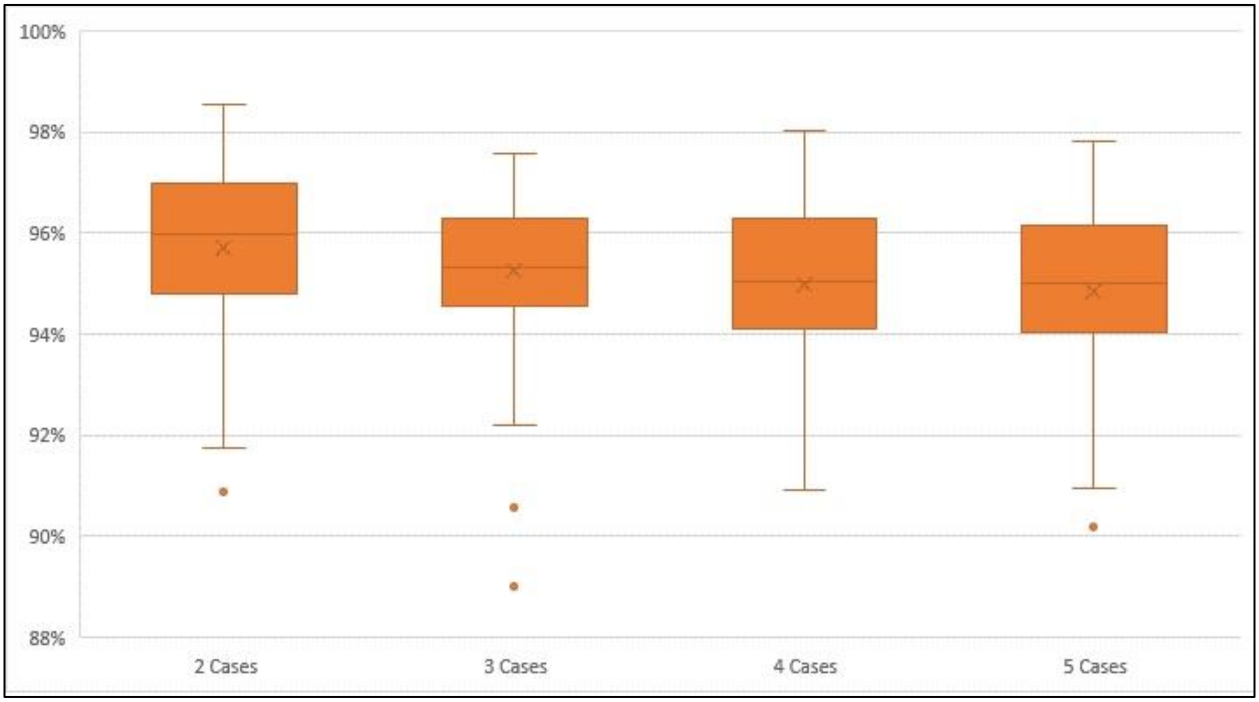
Figure 20 Coefficient of Determination for Multiple Billing Code Linear Regressions



Data from Vizient Procedural Analytics used with permission of Vizient, Inc. All rights reserved.

Each multiple billing code simulation also had a ratio that performed better than baseline.

Figure 21 Best Performing End Times Compared to Baseline, All Multiple Billing Code Simulations

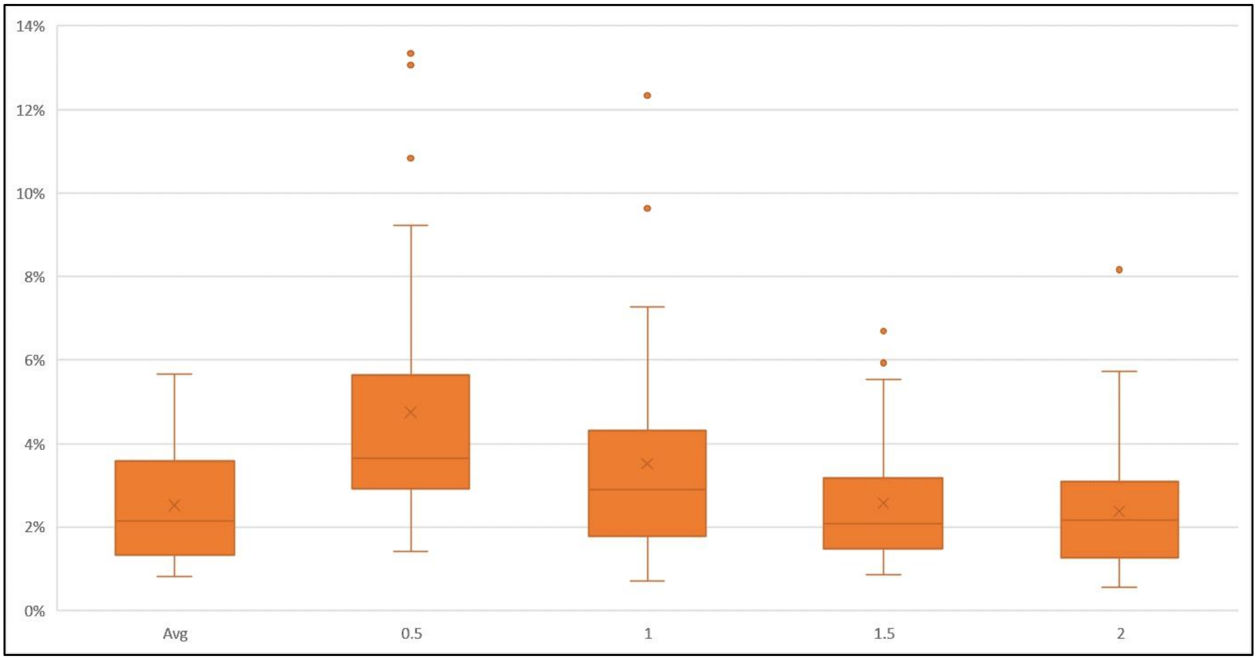


Data from Vizient Procedural Analytics used with permission of Vizient, Inc. All rights reserved.

Impact of Case Order

The last simulation looked at case ordering impacting delays and final case end times. A total of 37 physicians were randomly selected, and each schedule started with three different billing codes from all valid cases that physician performed. Each of 6 possible orders were simulated using schedules created by the average case length and ratios .5, 1, 1.5, and 2. The difference between the lowest and highest result for each order in each scheduling type was calculated for the case end time.

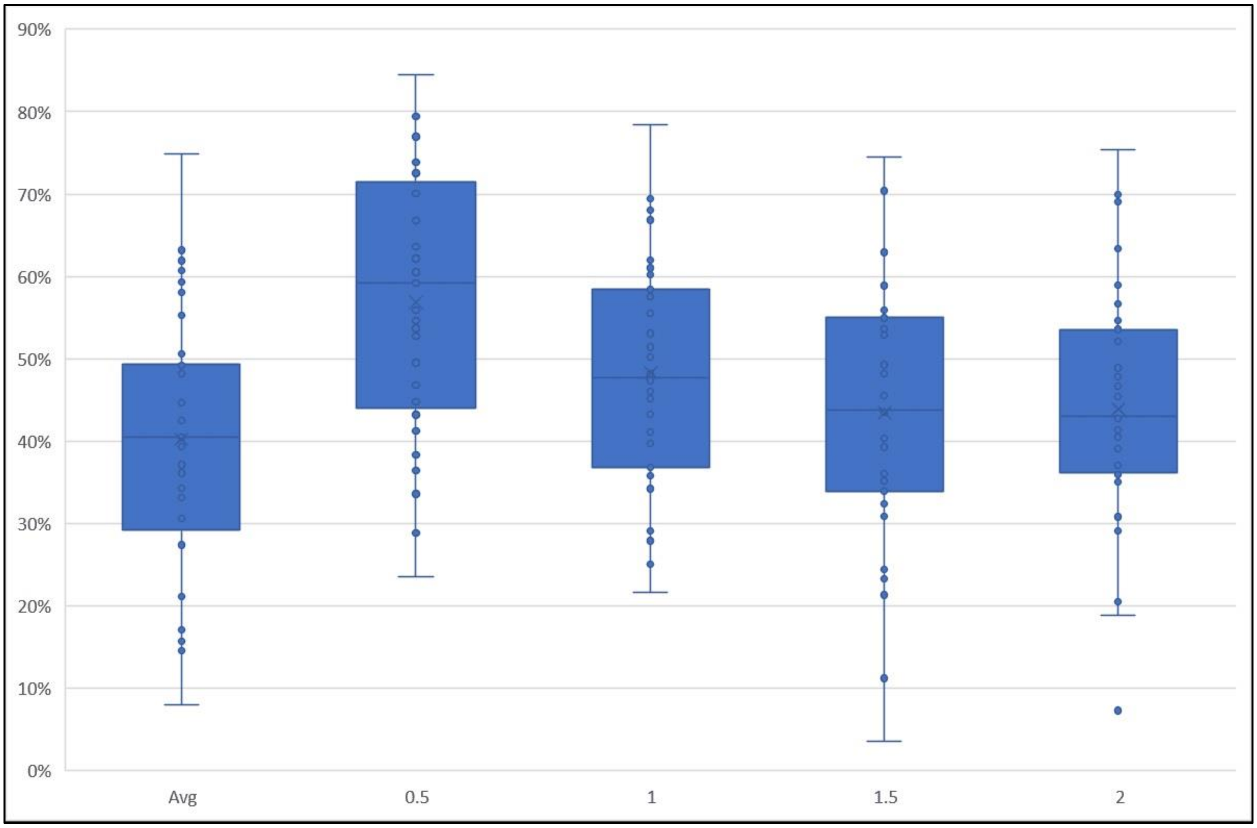
Figure 22 Percent Difference of Shortest and Longest Average End Times for Reordered Cases Over Longest Avg End by Ratio



Data from Vizient Procedural Analytics used with permission of Vizient, Inc. All rights reserved.

The same method was used to compare total delays. Unlike the time of the last case finishing, the smallest total delay was often close to zero, leading to larger ranges in the differences.

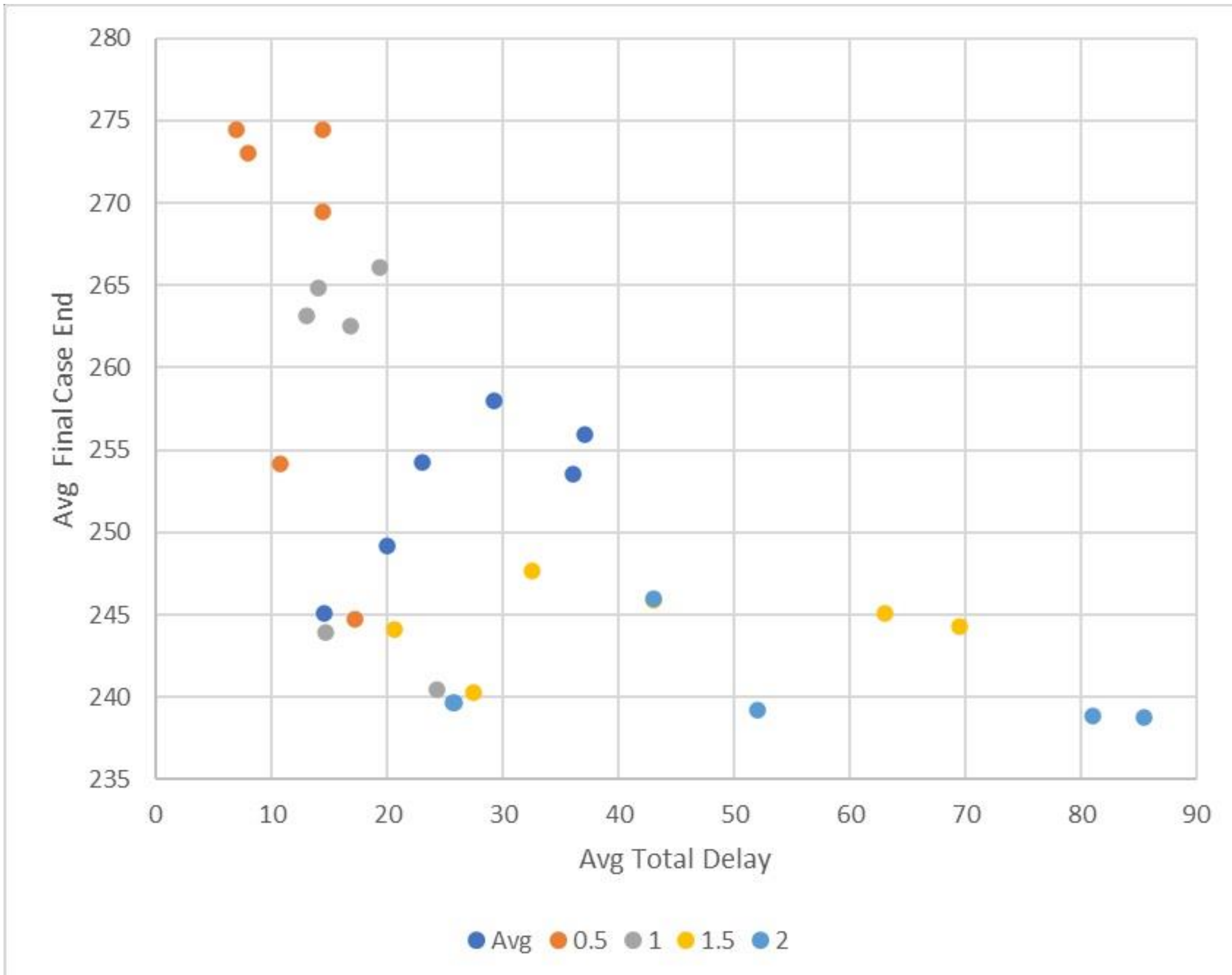
Figure 23 Percent Difference of Least and Most Case Delays for Reordered Cases Over Most Case Delays by Ratio



Data from Vizient Procedural Analytics used with permission of Vizient, Inc. All rights reserved.

Figure 24 visually shows all six order possibilities for one set of physician-billing code simulations with four weighing ratios and the baseline average scheduling method, and Appendix A: Additional Case Re-Ordering Results lists three additional combinations by ratio, order, and their resulting metrics.

Figure 24 Three Reordered Cases by Total Delay and Avg Final Case End with Five Different Scheduling Methods



Data from Vizient Procedural Analytics used with permission of Vizient, Inc. All rights reserved.

Chapter 5- Analysis & Conclusion

Scheduling by Weighing Compared to Case Average

Based on Figure 12 All Single Billing Code Simulation Results Compared to Baseline and Figure 18 Ratio of Total Case Delays over Baseline, All Multiple Billing Code Simulations by Ratio, the weighing ratios between .9 and 2 performed similarly to the baseline schedules. That range of ratios had performing better in both metrics, either metric, or neither metric. Scheduling with weights less than .9 consistently had less delays, and weights over 2 consistently were done quicker than scheduling by average case lengths. While it's possible any schedule could have a weighing ratio that creates a schedule that can routinely outperform the corresponding baseline schedule, the improvements in both metrics would be small percentages based on Figure 18 Ratio of Total Case Delays over Baseline, All Multiple Billing Code Simulations by Ratio and Figure 19 Ratio of Final Case End Times Compared to Baseline, All Multiple Billing Code Simulations by Ratio. Still, small improvements from only a subtle administrative change with no staff retraining or purchasing of equipment that can be adjusted based on each individual day's case types is something to continue pursuing. The scheduling-focused models discussed in the Modeling and Simulations section also had small improvements.

Additionally, using the average case length also limits the ability to tweak the schedule. Each simulated scenario focused only on randomly chosen cases and ignoring the rest of how a hospital interacts with a patient. Figure 10 Weighted Simulated Total Knee Replacement Case Timeliness with Ratio 1.6 By Order In Day and Table 15 Scheduled Case Lengths as Percentile of Cases in Billing Code Simulations show that there is a way to adjust a schedule by changing the scheduled case length based on when it occurs in the day to improve the timeliness of later cases. An Initial Simulation showed how scheduling cases by only the average case length leads to additional delays as the day progresses. More cases scheduled by average means more cases being delayed as the day progresses.

Case Ordering

Using only three cases per reordering simulation may have allowed for 20 times as many different simulations compared to using five cases as in the other simulations, but it also may have dampened the effects of reordering. Figure 24 shows why more investigation would be useful, as the benefits of trimming 15 to 25 minutes of around 250 minutes of operating time at ratios .5 and 1 is a time savings worth pursuing. But the entries Appendix A: Additional Case Re-Ordering Results show that there are other case combinations that saw less time savings in longer operating room days. The question started off as "Can case reordering reduce total time spent in the operating room?" After these results, it should now be “What makes case reordering more effective in some sets of cases over others?”

Future Steps

To focus the simulation results on the differences in scheduling methods, the metrics created were designed to compare cases with wildly different case lengths. In real operating rooms, non-emergency cases are scheduled by available blocks of time based on staffing. There are two questions that this modeling can help answer.

1. What is the likelihood of finishing this selection of cases within a given block of time?
2. If the cases go longer than that block of time, by how much?

The answers to those questions depend on the normal staffing blocks of time as well as how overtime is paid, which are hospital specific variables. But the idea of the variance of the last case’s ending time can be studied by the data that was made available for this study. One of the simulation rules stipulated that no case could start before it is scheduled to start. That results in any case performed under the scheduled time becoming as long as the scheduled time for that simulated day. The effect of picking the larger of the case scheduled length and the actual case length should mean that the average length in the schedule increases while the variance decreases. But, scheduling to reduce total work time would increase the variance as the delays could accumulate, as that maximum of scheduled case length and actual case length becomes the actual case length when the schedule case lengths decrease.

Adding in turnaround time between cases would help make the simulations more applicable to real schedules. Like the length of time cases can be scheduled, different hospitals may have different rules-of-thumb for including turnaround time in the schedule and different processes for cleaning and preparing an operating room. This addition would make for meaningful simulations for flip rooms, where one physician’s cases are schedule in two rooms with the idea that the cut to close time in one room occurs while the other room is turned over and the patient is prepped. Being able to simulate the effectiveness of room flipping would help determine which combinations of cases make flipping worthwhile enough to dedicate additional resources to one physician over another.

Closely related to flipping is studying when to schedule a single case closer to the end of the day. The discussion around turnaround time found out there's a strong preference for a surgeon to follow themselves in a room. For the times that doesn't occur, the different surgeon may not be picky about which room they perform the case in, just that they know when it occurs. If there are three possible rooms that all have some cases already scheduled, simulating those rooms to find the earliest time any of those three rooms ends is when that case can be scheduled instead of picking one of the three rooms at the start of the day. Combining the concepts, if all remaining scheduled cases are to be performed by physicians already operating, simulations can help decide which physician should flip into an available room to remove some downtime and reduce staff overtime.

The linear regression equations applied to the results strongly suggest more work could be done to reduce the number of simulations needed. Simulating multiple weighing ratios for a schedule to see how long the cases would take and how delayed they'd be isn't feasible when working with many rooms or reordering cases. Quicker ways to estimate those results would make scheduling software easier to use. The simple simulations used one case five times, making using that's cases mean and variance in a linear equation easy, but the reordering simulations showed different cases in different orders had different results. That strongly suggests the ordering of those means and variances for different cases would be meaningful in revisions to the linear regressions.

Finally, grouping billing codes that are similar in both medical terms and statistical analysis of their case lengths would help in two ways. First, some of the billing codes that weren't used often could be added to a more frequent code to allow them to be simulated in a meaningful way. Second, hospitals may not schedule by billing code. The 21 different codes for an appendectomy may just be scheduled as an appendectomy, and a medical coder reviews the case ones to select the appropriate billing code. Finding out the appropriate billing code after the surgery was performed doesn't help in scheduling that case at all.

Conclusion

Using average case lengths to schedule operating room case lengths doesn't allow room for alterations based on the other cases. This paper demonstrates simulation methods that look at how to schedule an operating room based on how prior cases influence when later cases start. Scheduling by simulation resulted in either days finishing quicker or reduced delays depending on how the empty operating room and case delay minutes were each weighed. There were some ratios where the simulated schedules performed very closely with the corresponding scheduling by average. As a current process though, scheduling by simulation requires too much computing time to be used. There is evidence that a linear regression equation could be developed to help guide what ratios to use based on the resulting metrics. How the coefficients for the variables in that linear regression equation would be set would depend on the ordering of the cases. For most simulations, reordering cases had a minimal impact on time saved in the operating room, but there were enough with enough time savings to warrant further investigation.

Overall, more research needs to be done on trends on underlying case variance impacting scheduling before implementing this method at hospitals. There was not much found in how different cases with different averages and variances impacted the schedule, only in single repeated cases. Even with small and occasional reductions in total operating day lengths found, the demonstration of balancing delayed cases throughout the day makes pursuing this scheduling method worthwhile over the trend of scheduling by average lengths meaning more delays as time passes.

Bibliography

Abbou, B., Tal, O., Frenkel, G., Rubin, R., & Rappoport, N. (2022). Optimizing operation room utilization - a prediction model. *Big Data and Cognative Computing*, 76.

Advisory Board. (2022, March 7). *How 7 health systems are tackling delayed surgery backlogs* . Retrieved from Advisory Board: https://www.advisory.com/daily-briefing/2022/03/07/delayed-surgeries

Alles, D., & Burshek, J. (2016, June). *Ranking analytics maturity by industry*. Retrieved from International Institute for Analytics: https://iianalytics.com/resources/ranking-analytics-maturity-by-industry

Ayres, C., & Trout, A. (2022, April 29). *Why focus on operating room prime time utilization: a metric for better use of OR time*. Retrieved from LeanTaas: https://leantaas.com/blog/optimizing-your-operating-rooms-prime-time-utilization/

Betts, D., Balan-Cohen, A., Shukla, M., & Kumar, N. (2016, July 8). *The value of patient experience.* Retrieved from Deloitte US | Audit, Consulting, Advisory, and Tax Services: https://www2.deloitte.com/content/dam/Deloitte/us/Documents/life-sciences-health-care/us-dchs-the-value-of-patient-experience.pdf

Bulmer, M. G. (1979). Properties of Distributions. In M. G. Bulmer, *Principles of Statistics* (pp. 62-63). New York: Dover Publications.

CaseCtrl. (2021, June 4). *5 of the most important operating room performance metrics*. Retrieved from CaseCtrl: https://blog.casectrl.com/5-of-the-most-important-operating-room-performance-metrics

Cerfolio, R. J., Ferrari-Light, D., Ren-Fielding, C., Fielding, G., Perry, N., Rabinovich, A., . . . Patcher, H. L. (2019). Improving operating room turnover time in a New York City academic hospital via lean. *Annals of Thoracic Surgery*, 1011-1016.

Cheng, H., Clymer, J. W., Po-Han Chen, B., Sadeghirad, B., Ferko, N. C., Cameron, C. G., & Hinoul, P. (2018). Prolonged operative duration is associated with complicates: A systematic review and meta-analysis. *The Journal of Surgical Research*, 134-144.

Colosi, B. (2022). *2022 NSI National Health Care Retention & RN Staffing Report.* East Petersburg, PA: Nursing Solutions Inc.

Common Ground Editorial Board. (2021, July 4). *Accurate case scheduling*. Retrieved from Salem Health: https://www.salemhealth.org/for-healthcare-professionals/common-ground/common-ground-2021/july-4-2021/accurate-case-scheduling

Foster, T. (2012, January). *Data for benchmarking your OR's performance.* Retrieved from OR Manager: https://www.ormanager.com/wp-content/uploads/2012/01/0112_ORM_5.Benchmark_r.pdf

Fullerton, T., Kore, A., Ninan, B., Ninan, D., Wasson, E., & Zhu, J. (2017). The role of organization culture in operating room turnaround time. *Cureus*, e1257.

Guerriero, F., & Guido, R. (2011). Operational research in the management of the operating theater: A survey. *Health Care Management Science*, 89-114.

Hatton, R. (2022, March 10). *Equipment, property cost $1.4M per operating room at ASCs*. Retrieved from Becker's Healthcare: https://www.beckersasc.com/benchmarking/equipment-property-cost-1-4m-per-operating-room-at-ascs.html

Haverfield, M. C., Tierney, A., Schwartz, R., Bass, M. B., Brown-Johnson, C., Zionts, D. L., . . . Zulman, D. M. (2020). Can patient–provider interpersonal interventions achieve the quadruple aim of healthcare? A systematic review. *Journal of General Internal Medicine*, 2107-2117.

Hicks, K. B., Glaser, K., Scott, C., Sparks, D., & McHenry, C. R. (2020). Enumerating the causes and burden of first case operating room delays. *The American Journal of Surgery*, 486-489.

Jones, D. (2018, March 3). *ICD-10-CM, ICD-10-PCS, CPT, and HCPCS: What are they?* Retrieved from Medical Coding Buff: https://www.medicalcodingbuff.com/icd-10-cm-icd-10-pcs-cpt-and-hcpcs/

Li, F., Gupta, D., & Potthoff, S. (2016). Improving operating room schedules. *Health Care Management Science*, 261-278.

Mayo Clinic Staff. (2022, December 18). *Breast reconstruction with implants*. Retrieved from Mayo Clinic: https://www.mayoclinic.org/tests-procedures/breast-reconstruction-implants/about/pac-20384934

McIsaac, D. I., Abdulla, K., Yang, H., Sundaresan, S., Doering, P., Vaswani, S. G., . . . Forster, A. J. (2017). Association of delay of urgent or emergency surgery with mortality and use of health care resources: A propensity score–matched observational cohort study. *Canadian Medical Association Journal*, E905-E912.

Obermair, A. (2016, November 16). *Why is there sometimes a wait at the hospital before surgery? Professor Andreas Obermair*. Retrieved from Professor Andreas Obermair: https://www.obermair.info/latest-news/blog/surgical-wait-time/

Oussedik, S., MacIntyre, S., Gray, J., McMeekin, P., Clement, N. D., & Deehan, D. J. (2021). Elective orthopaedic cancellations due to the COVID-19 pandemic: where are we now, and where are we heading? *Bone & Joint Open*, 103-110.

Pappada, S. M., Papadimos, T. J., Khuder, S., Mack, S. T., Beachy, P. Z., & Casabianca, A. B. (2022). Contributing factors to operating room delays identified from an electronic health record: A retrospective study. *Anesthesiology Research and Practice*, 1-7.

Peters, J., Basham, L., & Cole, A. (2021, April 7). *How to rebuild surgical revenue after COVID-19 (even if you just lost 60% of your OR volume)*. Retrieved from Becker's Hospital Reveue:

https://www.beckershospitalreview.com/strategy/how-to-rebuild-surgical-revenue-after-covid-19-even-if-you-just-lost-60-of-your-or-volume.html

Rahimi, I., & Gandomi, A. (2020). A comprehensive review and analysis of operating room and surgery scheduling. *Archives of Computational Methods in Engineering*, 1667-1688.

Rozario, N., & Rozario, D. (2020). Can machine learning optimize the efficiency of the operating room in the era of COVID-19? *Canadian Journal of Surgery*, E527-E529.

Sadler, D. (2016, September 21). *Building blocks: How block scheduling promotes efficiency*. Retrieved from OR Today: https://ortoday.com/building-blocks-how-block-scheduling-promotes-efficiency/

Slyter, K. (2018, November 19). *Types of hospitals: Your go-to guide for deciphering the differences*. Retrieved from Rasmussen University: https://www.rasmussen.edu/degrees/health-sciences/blog/types-of-hospitals/

Sokal, S. M., Craft, D. L., Chang, Y., Sandberg, W. S., & Berger, D. L. (2006). Maximizing operating room and recovery room capacity in an era of constrained resources. *Archives of Surgery*, 389-395.

The Cleveland Clinic Foundation. (2018, April 10). *Diabetes guidelines before surgery.* Retrieved from Cleveland Clinic: https://my.clevelandclinic.org/-/scassets/files/org/patients-visitors/information/prepare-for-surgery/2018-04-10-diabetes.ashx?la=en

Vizient Inc. (2022). *Procedural Analytics*. Retrieved from Vizient: https://www.vizientinc.com/what-we-do/supply-chain/procedural-analytics

Vizient, Inc. (2023, March 9). Vizient Procedural Analytics. Irving, TX, United States of America: Vizient, Inc. Retrieved March 8, 2023, from https://www.vizientinc.com

Wu, A., Huang, C.-C., Weaver, M. J., & Urman, R. D. (2016). Use of Historical Surgical Times to Predict Duration of Primary Total Knee Arthroplasty. *The Journal of Arthroplasty*, 2768-2772.

Wu, A., Kodali, B. S., Flanagan, H. L., & Urman, R. D. (2017). Introduction of a new electronic medical record system has mixed effects on first surgical case efficiency metrics. *Journal of Clinical Monitoring and Computing*, 1073-1079.

Yount, K. W., Lau, C. L., Yarboro, L. T., Ghanta, R. K., Kron, I. L., Kern, J. A., & Ailawadi, G. M. (2015). Late operating room start times impact mortality and cost for nonemergent cardiac surgery. *The Annals of Thoracic Surgery*, 1653-1659.

Zhu, S., Fan, W., Yang, S., Pei, J., & Pardalos, P. M. (2019). Operating room planning and surgical case scheduling: A review of literature. *Journal of Combinatorial Optimization*, 757-805.

Appendices

Appendix A: Additional Case Re-Ordering Results

Physician – Billing Code Combinations	Weighing Ratio	Case Order	Average Total Delay	Average Final Case Ending
A	0.5	1	11.971	306.922
A	0.5	2	5.26	303.124
A	0.5	3	15.397	308.168
A	0.5	4	15.815	303.991
A	0.5	5	6.169	300.206
A	0.5	6	12.758	301.234
A	1	1	15.353	301.573
A	1	2	8.713	298.362
A	1	3	22.364	301.042
A	1	4	21.912	297.055

A	1	5	8.744	296.092
A	1	6	17.117	292.854
A	1.5	1	21.068	299.383
A	1.5	2	12.763	294.416
A	1.5	3	28.634	295.032
A	1.5	4	27.527	293.161
A	1.5	5	14.205	295.533
A	1.5	6	19.087	293.468
A	2	1	23.919	296.76
A	2	2	15.693	294.09
A	2	3	28.943	296.121
A	2	4	29.614	294.531
A	2	5	13.749	292.712
A	2	6	20.663	294.331
A	Avg	1	15.798	300.623
A	Avg	2	11.358	295.859
A	Avg	3	20.759	302.316
A	Avg	4	18.134	300.117
A	Avg	5	10.481	297.209
A	Avg	6	14.04	300.789
B	0.5	1	15.958	192.269
B	0.5	2	21.281	191.245
B	0.5	3	9.849	194.873
B	0.5	4	10.153	190.41
B	0.5	5	13.334	194.403
B	0.5	6	10.66	190.707
B	1	1	22.558	184.541
B	1	2	30.025	182.934

B	1	3	17.168	182.957
B	1	4	15.826	183.165
B	1	5	23.031	183.236
B	1	6	14.088	185.772
B	1.5	1	26.495	178.765
B	1.5	2	30.707	178.68
B	1.5	3	21.076	180.206
B	1.5	4	15.557	182.99
B	1.5	5	24.358	181.695
B	1.5	6	17.902	184.215
B	2	1	30.285	181.896
B	2	2	31.564	178.985
B	2	3	21.791	180.959
B	2	4	18.749	181.842
B	2	5	23.206	181.551
B	2	6	19.529	183.829
B	Avg	1	20.111	187.279
B	Avg	2	21.434	188.998
B	Avg	3	15.791	185.41
B	Avg	4	14.088	184.141
B	Avg	5	17.925	189.117
B	Avg	6	15.24	184.441
C	0.5	1	8.318	169.301
C	0.5	2	9.887	168.504
C	0.5	3	9.818	174.553
C	0.5	4	13.113	178.565
C	0.5	5	12.945	170.253
C	0.5	6	18.872	171.001

C	1	1	11.137	165.64
C	1	2	12.901	166.547
C	1	3	17.672	166.385
C	1	4	22.924	166.833
C	1	5	18.917	165.667
C	1	6	21.841	166.521
C	1.5	1	12.091	163.567
C	1.5	2	13.393	165.003
C	1.5	3	19.105	162.068
C	1.5	4	25.665	163.817
C	1.5	5	19.425	161.591
C	1.5	6	22.469	163.044
C	2	1	14.562	163.139
C	2	2	14.528	162.461
C	2	3	20.352	163.296
C	2	4	26.469	164.082
C	2	5	22.792	162.274
C	2	6	30.352	162.989
C	Avg	1	9.719	166.249
C	Avg	2	10.507	166.076
C	Avg	3	14.759	167.363
C	Avg	4	18.249	171.488
C	Avg	5	14.621	166.651
C	Avg	6	19.121	171