Improving Healthcare and Operating Room Efficiency Using Lean Six Sigma

by

Sarah Slagerman

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfilment of the requirements of the degree of

MASTER OF SCIENCE

Department of Biosystems Engineering

Faculty of Engineering

University of Manitoba

Winnipeg

ABSTRACT

Wait times and availability of care are major issues within Manitoba's healthcare system and to improve in these areas a new method for improvement is needed. The purpose of this research is to prove the efficacy of using Lean Six Sigma in healthcare to generate improvements and to promote the usage of continuous improvement methodologies in the healthcare environment.

To demonstrate the effectiveness of using Lean Six Sigma, a project was completed at St. Boniface Hospital to reduce overtime in operating rooms. Lean Six Sigma was used to assess the entire system and identify multiple areas for improvement, with case duration estimates being found to have the most potential for reducing overtime. This resulted in predictive models being created and tested against the current method of surgeon estimates. All models improved on the surgeon estimates (45-63% increase in on-time cases, reduction in overtime error by 49-59%, and 71-89% improvement in overtime to undertime error ratio) and it is recommended that a predictive modeling approach be used in the future.

The Lean Six Sigma project was successful and also resulted in multiple additional beneficial outcomes: identification of other areas needing improvement, ranked by potential impact; process analysis and mapping which can be used in future projects; and identification of other causes for error in scheduling. In addition, if Lean Six Sigma had not been used, the project would have focused on a less impactful area—first case on-time starts. As Lean Six Sigma is a data-driven process, the impact of bias was removed and thus it was found that first case on-time starts were not as influential to overtime as assumed. From this research, it can be concluded that Lean Six Sigma can be effectively applied in even the most complex of hospital environments. It is recommended that hospitals consider implementing experienced teams to lead and train hospital employees in Lean Six Sigma or other continuous improvement methods.

ACKNOWLEDGEMENTS

I would like to take this opportunity to acknowledge the many people who supported and guided me throughout my research.

- My advisor, Dr. D. Mann, for your guidance and leniency in accepting a grad student who began without a solid plan for a thesis topic
- Dr. D. Petkau and Dr. Qingjin Peng, for being a part of my advisory committee
- Dr. J. Morrison, for teaching me how to use R and then helping with the seemingly endless questions which stemmed from that learning
- Vern Campbell, for introducing me to Op Ex and to individuals at St. Boniface Hospital
- Kurt Shaw, for listening to my vague dreams for my thesis and helping me get into a project for which I am passionate
- Lance Barber, Dawn Affleck, Lorena Thiesson, and Tamara Miller, for being a part of my executive team at St. Boniface and providing me with resources and support
- Mohammed Yusuf and Scott Vandale, for your help with getting me the thousands of lines worth of data which I needed for this project
- Jennifer Mills, for your help with scheduling (and rescheduling) all of our meetings
- St. Boniface Transformation Team, for your help, guidance, and information
- To all the St. Boniface team and staff members who allowed me to shadow them, thank you for putting up with me and patiently answering my many questions
- Province of Manitoba and UofM Faculty of Graduate Studies, for financial assistance provided through the Manitoba Graduate Scholarship and NSERC-CGSM, respectively

DEDICATION

This thesis is dedicated to my mother and father. The example you have set for me and the support you provide is everything. Also, to Curtis: thanks for pushing me to get this done.

GLOSSARY OF TERMS/ABBREVIATIONS

Improvement Methods

Lean:	continuous improvement process focusing on reducing and eliminating waste
Six Sigma:	continuous improvement process focusing on reducing cycle time and variation
OpEx:	operational excellence; company-wide mindset of continuous improvement
VMPS:	Virginia Mason Production System
MCQC:	Mayo Clinic Quality Construct
CCIM:	Cleveland Clinic Improvement Model
TIS:	ThedaCare Improvement System

Lean Six Sigma Terminology

CTQ:	customer critical-to-quality characteristics
SIPOC:	supplier/inputs/process/outputs/customer
DMAIC:	define/measure/analyse/improve/control
PDC(S)A:	plan/do/check(study)/act
VSM:	value stream mapping
RIE:	rapid improvement event
DPMO:	defects per million opportunities
DPU:	defects per unit

Operating Room Terminology

OR:	operating room
L2PO:	pre-operative areas
Slating:	scheduling method for surgeries
HCA:	health care aide
ASA:	American Society of Anesthesia
Overtime:	running greater than 15 minutes over schedule
On-time:	finishing within 15 minutes of the scheduled time/duration
Undertime:	running greater than 15 minutes under schedule

Surgical Procedures

CABG: coronary artery bypass graft

AVR: atrial valve replacement/repair

MVR: mitral valve replacement/repair

END: carotid endarterectomy

HAT: total abdominal hysterectomy

HYSD: diagnostic hysteroscopy

Predictive Models

SMA: simple moving average

LME: linear mixed effects

RF: random forest

Surgery Parts/Durations

Preparation (PREP): duration from patient entering the OR to anesthesia being induced

Anesthesia (ANES): duration from anesthesia being induced to incision

Procedure (PROC): duration from incision to dressing ("cut to close")

Wrap-up (WRAP): duration from dressing to the patient exiting the OR

Case (CASE): duration from the patient entering to exiting the OR ("toes in to toes out")
(preparation + anesthesia + procedure + wrap-up)

Turnover time: time between one patient exiting the OR and the next patient entering

Turnaround time: time between patient dressing to next patient incision
(wrap-up + turnover + preparation + anesthesia)

Room duration: duration from first patient entering to last patient exiting the OR

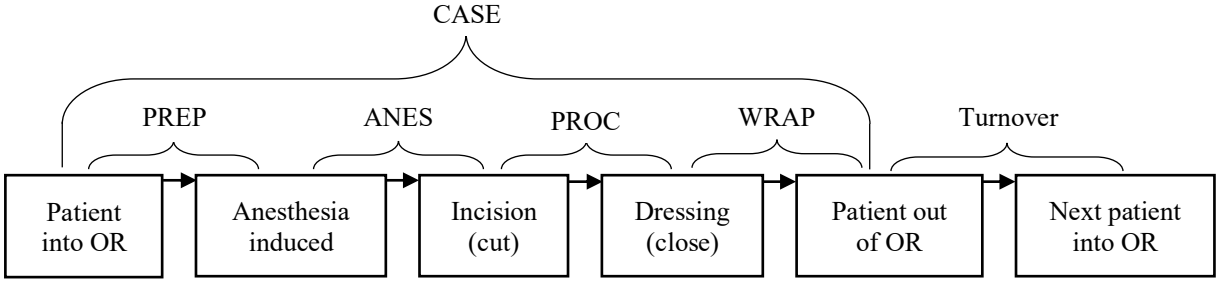


TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

This thesis follows the application of Lean Six Sigma at St. Boniface Hospital in a project to assess, prioritize, and improve areas for efficiency in the operating rooms (OR). With this work, the benefits of applying Lean Six Sigma in hospital and healthcare settings will be shown in order to aid in promoting the usage of continuous improvement philosophies throughout the healthcare sector in Manitoba. Along with the outcomes of the project, an assessment of barriers and complications with implementing this methodology in the hospital was completed.

1.1. Organization of Thesis

This thesis is separated into seven chapters, along with appendices. Chapter 1 introduces the research problem, as well as the importance and scope of the project. It also provides a background on Lean Six Sigma and outlines the steps used to implement the methodology. This is followed by a literature review in Chapter 2, which outlines variables that have been used to measure OR efficiency and methods used to improve healthcare efficiency. It also covers the shift in healthcare towards engineering and data-driven methods of improvement and how Lean Six Sigma fits into this role, along with challenges faced with implementation.

Chapter 3 outlines the project completed using the Lean Six Sigma methodology—reducing end-of-day late cases causing overtime by improving surgery duration estimates—and is broken into four sections: (3.1) define the initiative, (3.2) characterize the process, (3.3) improve, (3.4) monitor the outcomes. These are then further broken down into subsections based on the steps taken to complete each section. At the end of each of the first three chapters is a summary of key information covered.

Chapter 4 covers the results of applying Lean Six Sigma in the hospital. This includes a summary of outcomes generated by using Lean Six Sigma and how they may impact the

completion of future projects. This is then followed by a discussion in Chapter 5 on the effectiveness of using Lean Six Sigma to dissect and solve the problem and an overview of the challenges faced in implementing this methodology in the environment of St. Boniface Hospital. The final two chapters cover the conclusions from the project (Chapter 6) and recommendations moving forward (Chapter 7).

The appendices included in this thesis are used to explain elements used within the Lean Six Sigma project more thoroughly, including tools used for mapping the OR, details on the process used to select surgeries for further assessment, and data manipulation. The steps and coding used to create models for surgery duration predictions are also available for reference.

1.2. Research Problem and Scope

It is a fact of life in Canada that when a trip to the hospital is required, one does not get asked about if they have medical insurance to pay for what are often very expensive services; instead, they are provided the care that is needed. Indeed, many believe Canada to have one of the top healthcare systems in the world. However, in a comprehensive study which compared the healthcare systems of 11 high-income, developed countries [1], Canada ranked third-to-last, only achieving above France and the United States overall. The study groups the ranking metrics into five categories: care process, access, administrative efficiency, equity, and health care outcomes. Canada ranks in the bottom three in the areas of access, equity, and health care outcomes, and only sixth in care process and administrative efficiency. Some of the key areas where Canada fell short were in regard to wait-times for specialists and emergency room (ER) visits, poor after-hours availability for care, prevalence of common chronic conditions in the adult population, and the lack of coverage for dental work and prescription drugs.

Some may think that increasing the funding toward healthcare and reducing healthcare cuts would aid in reducing these problems. However, according to the Organization for Economic Co-operation and Development (OECD) [2], the health spending of Canada is higher per capita than some of the higher-ranking countries such as Australia and the United Kingdom, which were the two top ranked countries overall. In addition, the United States has the highest per capita spending and is ranked as the lowest performing healthcare system, thus showing that money alone will not provide results. A study done by the Fraser Institute [3] comparing only countries with universal healthcare systems corroborated Canada's issues with spending and shortfalls; it concluded that although Canada's healthcare is one of the most expensive universal healthcare systems in the OECD, availability and access to resources are below the average OECD country and there is room for improvement in the areas of resource usage and quality and clinical performance. It also stated that Canada ranked last on four of the five indicators of timeliness of care, proving once more the major issue that Canada has with wait times in healthcare.

Beyond the disappointment of comparing Canada's healthcare system with others in the world and having it fall short, there are the concerns for how these failings impact Canadians. As of 2019, the average waiting time for medically necessary elective surgery in Canada is 20.9 weeks, with Manitoba being above the average with a waiting time of 32.4 weeks from meeting with a general practitioner (GP) to treatment [4]. These wait times are of vast importance as it has been proven that longer wait times can have serious consequences, including loss of wages and economic costs [5] [6], undermining patient trust, extended pain and suffering, and in some cases poorer medical outcomes [7] [8]. However, as poignantly stated by André Picard from The Globe and Mail [9] on the issue of wait times for healthcare in Canada:

"… the real challenge we have is not waiting times. It is more fundamental: To provide the right care at the right place at the right time at an affordable price (to individuals and society)."

This thought perfectly describes the philosophy behind Lean Six Sigma, which is about getting the right parts (or in this case, care) to the right place, at the right time, in the right quantities while also reducing waste and error.

While these issues plague the Canadian healthcare system as a whole, Manitoba has waiting times longer than average within Canada. Thus, in order to reduce costs and improve the quality and timeliness of care, a structured method for improvement needs to be applied within our healthcare systems. That is why the objective of this research is to demonstrate the efficacy of using Lean Six Sigma to improve efficiency in healthcare and promote the application of Lean Six Sigma (and other continuous improvement methods) in healthcare services across Canada.

Changes in how we assess and improve efficiency need to be made on a national level, however they can start at a smaller scale such as within individual hospitals. Fortunately, the director of surgery at St. Boniface Hospital was interested in assessing and improving the efficiency of their OR and was open to the use of Lean Six Sigma to do so. The problem posed was the issue of operating rooms running overtime at end-of-day, resulting in uncoordinated patient flow and resource utilization, increased overtime hours and cancelled end-of-day cases, other staffing issues, and poor patient experience. As such, the Lean Six Sigma project objective was to establish recommendations on how to reduce the number of rooms running overtime from 125/month to 30/month.

In order to assess OR efficiency and recommend improvements, all processes required to run the OR for elective surgeries were considered, including, but not limited to, L2PO (pre-operative

area), OR room preparation, room turnover, start and end of day tasks, slating, and staffing or starting times. Areas that were outside the scope of this project included the elements or steps involved in the surgery or preparation itself, total work time for staff, and emergency surgeries. Project preparation began in February 2018, work and observation at the hospital began in April 2018, and on-site work was completed in October 2019. All data used complied with the Public Health and Information Act. This project did not require Research Ethics Board approval as the records were from past surgical procedures and could not be traced back to patients.

1.3. Intro to Lean Six Sigma

1.3.1. Quality Control, Continuous Improvement, and Operational Excellence

When it comes to process improvement, there are a few key terms used to define different types of improvement initiatives, namely: quality control, continuous improvement, and Operational Excellence (OpEx). The area of quality control revolves around ensuring a quality product which conforms to specification. If there are defects or faulty products, quality control is used to detect and prevent the errors. Continuous improvement deals with product quality as well, but also includes assessing and making improvements to all areas of production in order to reduce waste, cycle time, and costs. Any philosophy, concept, or methodology used to continually create improvements, either through incremental or breakthrough changes, is a continuous improvement method. OpEx goes a level above continuous improvement; it encompasses the mindset and ingrained culture of the company to being dedicated to creating constant and sustainable improvement through a continuous improvement system that utilizes problem-solving and management with set goals for the operational performance of the company. Depending on the goals set, a good OpEx method may also be a driving force for innovation and change within a company.

Many companies create their own unique methods for improvement from elements of existing methodologies in order to meet their own needs; these may also qualify as continuous improvement or OpEx methods. Lean Six Sigma is one example of a continuous improvement methodology that can be used within an OpEx management system.

1.3.2. Lean Six Sigma Development Timeline

The principles behind Lean Six Sigma are the combined elements of two different improvement methodologies: Lean manufacturing and Six Sigma. The important individuals and concepts which contributed to the development of these two methodologies are outlined briefly in the timelines below (Figure 1-1, Figure 1-2). For more information on the background of Lean Six Sigma, see APPENDIX A.

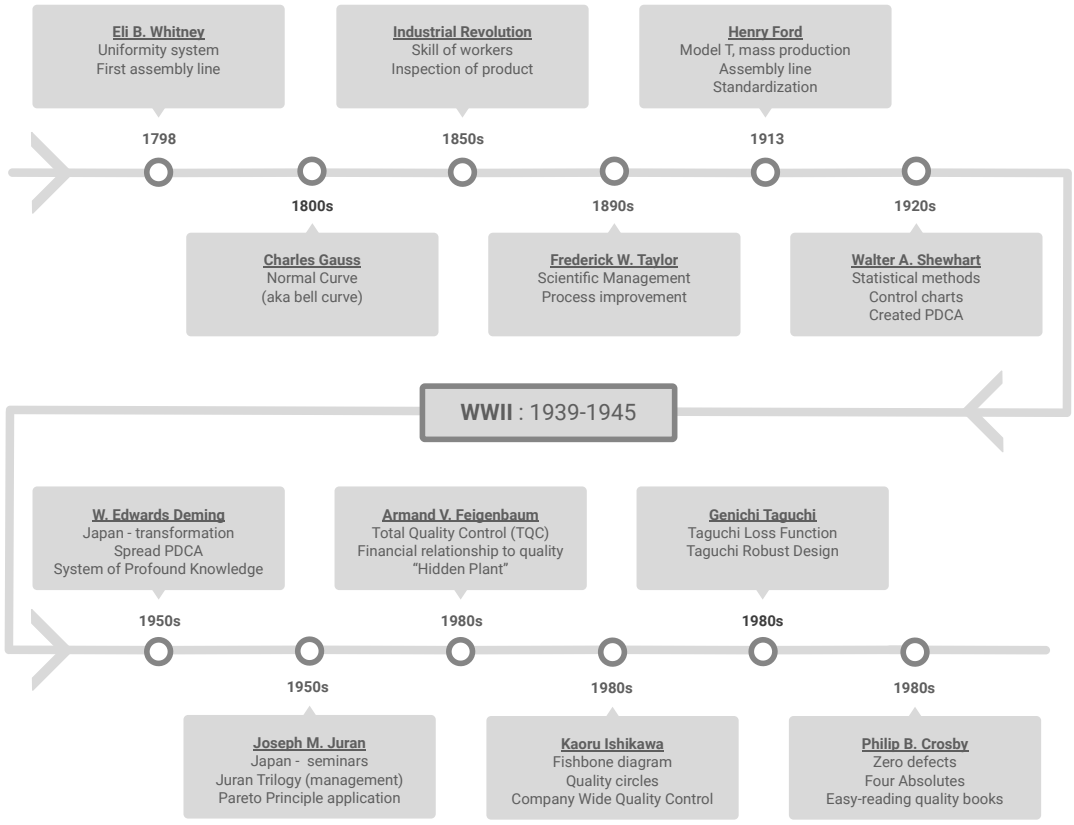


Figure 1-1 Quality Control and Efficiency Timeline

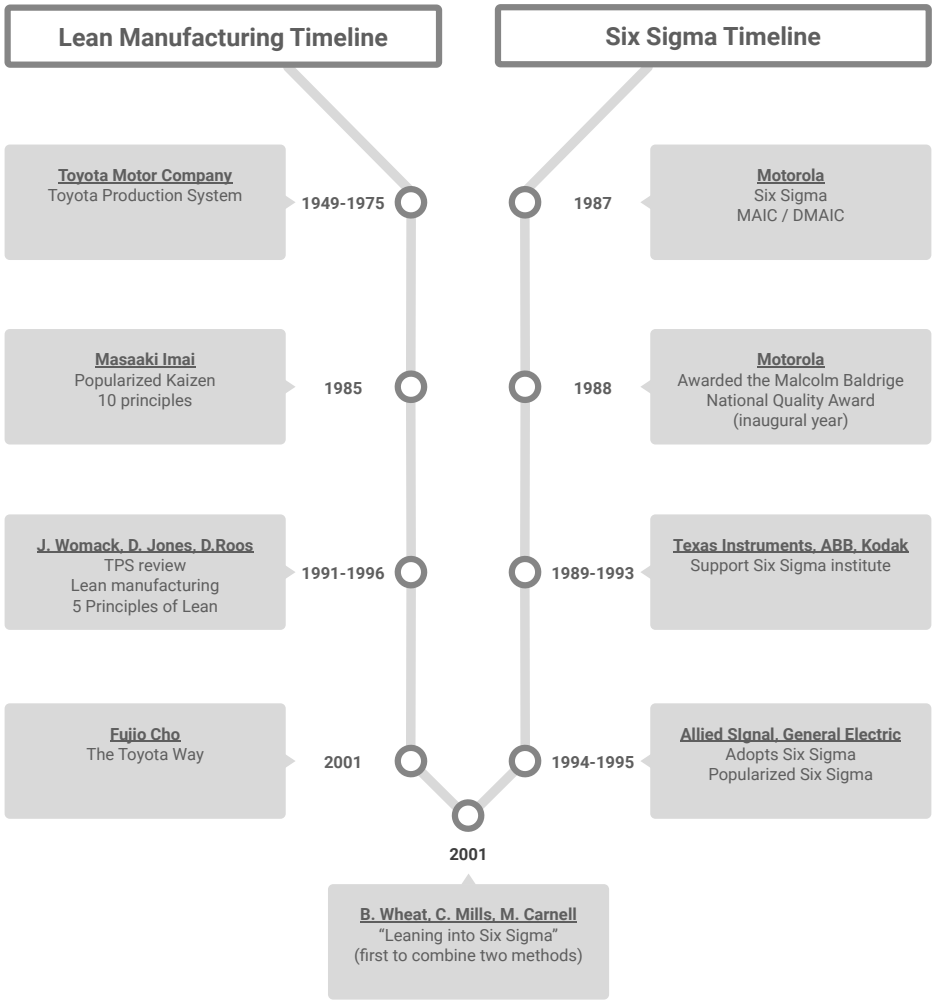


Figure 1-2 Lean Six Sigma Timeline

Lean focuses on reducing waste, improving process flow, and reducing cycle time. Six Sigma focuses on reducing defects, bringing processes under statistical control, and managing variation. Together, these two methods are able to address all areas of a process and allows for them to be assessed and improved to maximum effect.

1.3.3. Steps of Lean Six Sigma

Lean Six Sigma is a continuous improvement methodology through which problems within a process may be identified and solved systematically. The methodology is cyclical, so the process

is constantly being improved. For the purposes of this project, the Lean Six Sigma methods and steps used in the University of Manitoba engineering course *ENG7510 Operational Excellence* were followed. Figure 1-3 shows the steps undertaken when completing an improvement project. Each step is briefly explained within this section. The methodology will also follow this process.

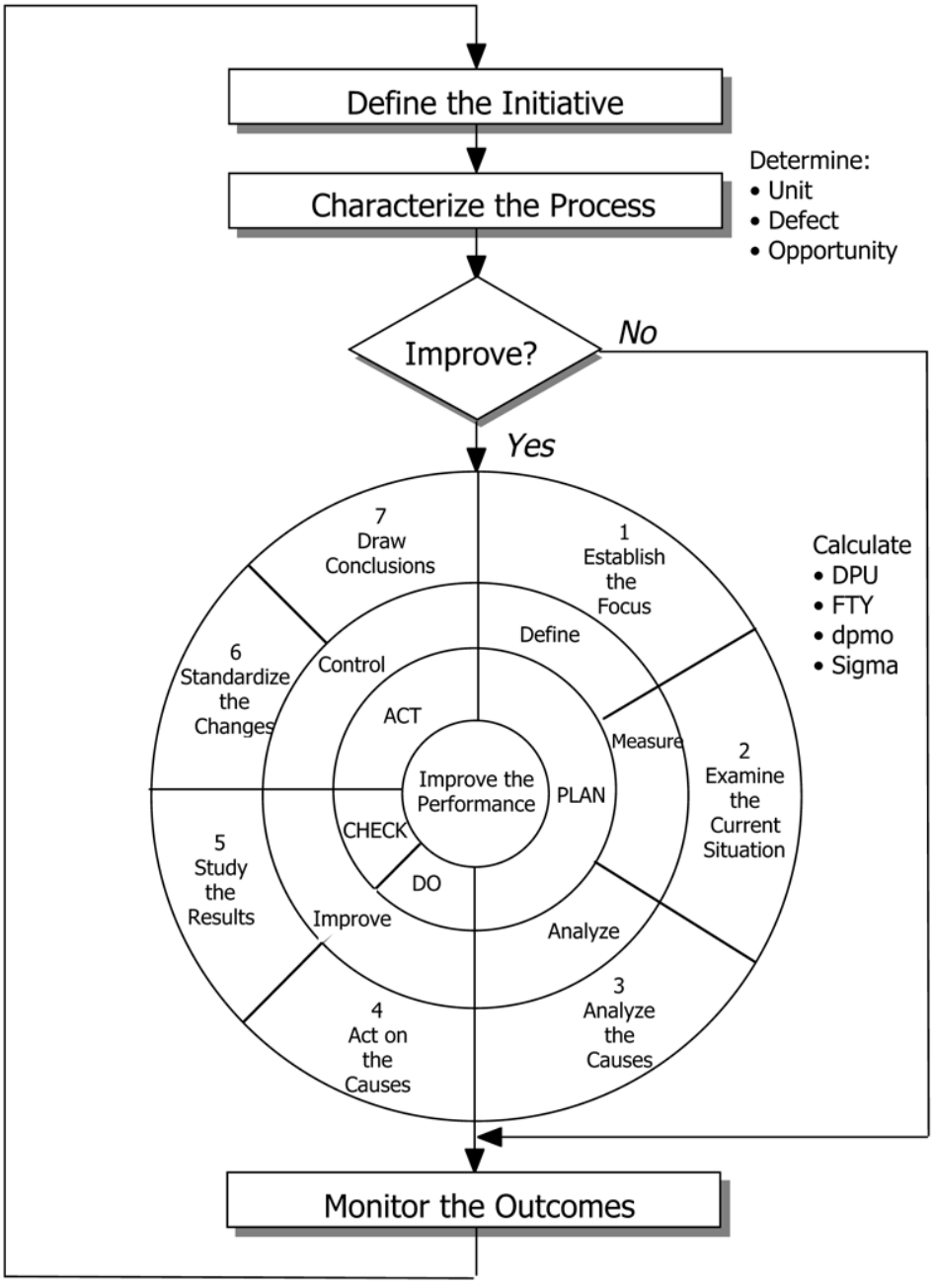


Figure 1-3 Lean Six Sigma Improvement Project [10]

1.3.3.1. Define the Initiative

To begin a Lean Six Sigma improvement project, one must first define the initiative. This involves writing a Project Charter and forming the project team. These elements ensure that the project has a clear, unified purpose and provides a framework for the team moving forward.

A Project Charter outlines many aspects of the project prior to work beginning and is a living document, so it may be updated and changed as necessary. It should clearly outline background, business case, objective, scope, schedule, roles and responsibilities, resources, and risks.

There are many roles and responsibilities associated with Lean Six Sigma projects. Identifying roles and responsibilities early ensures better communication and accountability of team members. The various roles often outlined in such projects can be seen in Table 1-1.

Table 1-1 Roles and Responsibilities [10]

Role	Responsibilities
Champion	Work with LSS expert and team leader to develop the Charter; attain resources and remove barriers to project; link between team and executives; advise/guide expert and leader; review progress
Lean Six Sigma Expert (Master/Black Belt)	Work with Champion and team leader to develop the Charter; educate/guide team on LSS methods; work with the team leader to lead the team; help with documentation/data analysis
Team Leader (Green Belt)	Work with Champion and LSS expert to develop the Charter; lead meetings; guide the team through the process; work as a team member, providing subject content and sharing the workload
Executive Management Team	Manage project portfolio; identify improvement initiatives; assign roles/responsibilities; connect team to financials/organization
Quality Department	Own improvement process and roll-out strategy; provide expertise, training, and support to organization
Process Owner	Accountable for processes, including making improvement and monitoring results; provides knowledge on how the process works (capabilities, outputs, connections)
Team Members	Follow leader/LSS expert direction; attend meetings; contribute subject content and share the project workload

For this project, the team roles and responsibilities are different from a normal Lean Six Sigma project in that the work was completed by only the team leader, with the other project members contributing feedback, guidance, and process knowledge. In addition, a Lean Six Sigma expert was not a part of the project. This was due to the project being completed in an

environment that was not currently applying Lean Six Sigma as a methodology for improvement. Instead, this project was done to prove the applicability and capability of the methodology through a test project.

1.3.3.2. Characterize/Understand the Process

This step in the Lean Six Sigma methodology is critical, especially for those who are new to a process or for a process which has not been assessed or improved using Lean Six Sigma previously. Characterizing the process allows the team to become familiar with the processes they plan to improve as well as the people who will be affected by the changes. It also provides a baseline to which changes can be measured. It is inadvisable to make changes to a system that one does not fully understand for it is unlikely to result in a positive change for all affected. In addition, this step provides a baseline for the current performance of a process. This step is broken into three sections to accomplish these requirements of understanding:

I. Identify process linkages (i.e., see how the process fits into the overall system)
II. Define how the process works (i.e., view the current workings and baseline performance)
III. Identify the performance issues (i.e., find areas for improvement)

In order to define the process, the customer critical to quality (CTQ) characteristics need to be identified and a SIPOC (supplier, input, process, output, customer) analysis completed. Once the performance issues have been identified, the next step is to improve.

1.3.3.3. Improve

Before beginning an improvement project, there must be adequate evidence to move ahead and garner support for the project. If a process is running at maximum capabilities and does not require improvement, skip to monitoring outcomes to ensure it sustains these values. Where possible, determine impact and prioritize potential projects. A budget and timeline may be

proposed for the potential projects at this time as well, to aid in determining which projects will be funded. Once one project has been agreed upon, the improvement process can begin. For this project, the DMAIC cycle (shown as the center ring in Figure 1-3) was used.

I. Define (Establish the Focus)

During this stage, the problem must be assessed in detail. The performance gap and proof of the problem (where the system is and where it should be performing) should be recorded and verified. Depending on the type of data, various different methods of data analysis may be utilized. At this time, a clear objective for the project and an improvement plan are also put in place. If needed, additional narrowing of the focus also occurs in this stage.

II. Measure (Examine the current situation)

In this stage of the improvement cycle, the process is assessed in greater detail specific to the area for improvement. This can be done using process maps, cause-and-effects matrix, capability analysis, or failure mode and effect analysis (FMEA). Not all of these tools are necessary for all projects; for this project, only process mapping and FMEA was completed. Any CTQs specific to the process should be given operational definitions and have baseline measurements taken. Control charts for the baseline measurements are used to assess sources of variation and to determine if the processes are stable. Once the baseline measurements are completed, strategies for meeting the project objective are determined, along with how to measure improvements.

III. Analyze (Analyze the causes)

In this stage, the goal is to determine the root causes of the current issues. To initially brainstorm these root causes, it may be beneficial to use fishbone diagrams, process knowledge and experience, and information from previous observation. After this, the link between the root

causes is discussed and verified (where possible) using data. Some tools which may be used to accomplish this include scatter plots, run charts, control charts, and pareto diagrams. Then, the key root causes must be selected for eliminating and the methods of measuring progress set. If there are multiple important issues which need to be addressed, a cost/benefit analysis on acting on each root causes may be completed to determine where to act first.

IV. Improve (Act on the causes, Study the results)

With the root cause for action chosen, brainstorming is done to create a list of possible actions. These are then assessed to determine which action will be taken forward. Depending on the constraints and requirements of the project, variables which may influence the action plan selection include cost, time, difficulty, impact, inter-connectivity issues with the rest of the process, and the level of authority needed to implement the chosen action. An action plan is then created and implemented on a smaller scale to test the improvement capabilities of the action.

The results of the small-scale action should be assessed using the previously set project measures. If the desired level of improvement is not achieved or if there are negative side effects to the action taken, then changes may be made to the plan and the outcome reassessed until the desired outcome is reached. Once the desired level of improvement has been accomplished, it is time to move to the next stage for full-scale implementation and standardization.

V. Control (Standardize the changes, Draw conclusions)

To standardize the successful changes from the small-scale implementation, a plan needs to be put in place on in order to ensure a smooth transition. This includes employee training as required (as well as changes to new employee training procedures), communication to all affected by the changes, and instigating methods for control and monitoring of the changes. Control and monitoring measures are particularly crucial in this step to ensure that the changes

are successful and sustained over the long-term; without them, the process may revert to its former state. Due to this project's timeframe and limited scope, the standardization of changes was not completed for this project. However, this may be completed in the future with additional resources focused on fitting the solution on a larger scale.

This is also the stage in which to reflect on and document the project. Reflection involves documenting any issues, lessons learned, and benefits from completing the project. The purpose of reflecting on the project in this way is to solidify knowledge on what was done during the improvement process in order to learn from mistakes and improve the process for future projects. Another element of this stage is to plan for the future—what future projects need to be completed and what needs to be done to get them started? This is where the cyclical nature of Lean Six Sigma comes into play as there should always be a next step, whether it is continuing with the current issue and tackling another of the root causes or moving to a new process and problem. Finally, the means of spreading the information and documenting the project must be decided. This is key to ensure that the information and analysis completed will be available for future works and also for spreading the knowledge of the changes throughout the organization.

1.3.3.4. Monitor the Outcomes

With the improvement portion of the Lean Six Sigma project completed, monitoring the outcomes as outlined under the "Control" step is a continuing necessity. This ensures that the project outcomes are reaching expectations and that there are no unexpected effects in the long-term. In addition, successful application of monitoring measures can allow for changes in the process to be picked up as time goes on and can inform when new improvements are needed. This step was not covered for this project as the changes were not fully implemented, they were only tested and thus do not require monitoring.

1.3.3.5. Repeat Cycle

Lean Six Sigma is a cyclical process, meaning that it will always be repeating and finding new areas for improvement. It is at this time that additional root causes may be identified for action within the same area, or a new project may be proposed.

1.4. Chapter Summary

In this introduction, the organization of the thesis was outlined (Section 1.1), along with details on the reasons for completing this work (Section 1.2). The research problem identified was the issue of efficiency in healthcare and the goal of this thesis is to promote the usage of continuous improvement methodologies such as Lean Six Sigma in the healthcare sector in Canada and prove they are effective. This is being accomplished through completing an improvement project using the Lean Six Sigma methodology to assess the OR at St. Boniface Hospital. As part of the introduction to the Lean Six Sigma methodology (Section 1.3), a timeline on how Lean Six Sigma was developed was included, along with a brief description of what each step in the improvement project entails. Additional information on the background of Lean Six Sigma is available in APPENDIX A.

2. LITERATURE REVIEW

No matter the venture, efficiency holds a vital role. In business, efficiency means profitability. In one's personal life, efficiency means having more time for whatever is important to them, be it work, family, fun, or rest. In healthcare, efficiency is especially important because it can save lives. A hospital that runs efficiently is able to optimize their resources—money, time, staff, medications, space—in order to treat the maximum number of patients possible. The more efficient our healthcare system is, the more value can be seen for the money input and with less effort.

The value for improving efficiency in healthcare may overarchingly be with the increase in productivity and thus the number of patients cared for and lives saved, however the benefits can be realized in many areas. For a hospital, the improvement in efficiency could result in more money being available to improve technology, hire staff, and maintain or update infrastructure. For staff, an improvement in efficiency may mean fewer steps taken in a day, more accurate work schedules, or less time spent waiting, for example, for equipment to become available. On scheduling, especially, the gains from efficiency may be seen as it could impact how many individuals would be willing to apply for certain positions; positions which traditionally require overtime would not be appealing for many (particularly those with young families or who are caregivers), thus resulting in fewer being interested in the positions and leaving even more of a requirement for overtime from those who are able. For patients, efficiency would reduce their wait times for surgery, consultation, and results as well as possibly improving their recovery rate and quality of care while in the hospital. Even for a patient's family, healthcare efficiency is important because it would allow for more accurate estimates of surgery durations, improve the ease of navigation through the hospital, and result in an overall improved experience.

While the benefits of an efficient healthcare system are clear, measuring and improving efficiency is more complex. For what makes a system efficient? In a complex system such as operating rooms, there are many areas and ways in which efficiency has been measured and improved, and the methods of improvement continue to evolve.

2.1. Measuring OR Efficiency

Depending on the area to which you look in the OR, efficiency may mean something different. As described by Vig et. al. [11], the measure of efficiency may be budgeting and throughput for administrative services, cancellation rate and first case on time starts for surgeons, patient injury rate for risk management, and still other measures for different departments such as nursing or anesthesia. Due to these differences in view on what makes an OR efficient, there are many variables against which efficiency may be measured.

This is only one of the issues faced when trying to measure efficiency in ORs. Other issues include lack of standard definitions for various OR processes, differences in methods used between studies, a lack of validation completed for indices used as OR performance indicators, and the difficulty of setting universal benchmarks between different hospitals due to variability in multiple areas. Hospitals may differ in patient population, type of hospital (teaching/for profit/not for profit), and type of surgery or anesthesia offered [12]. With these issues in place, it can be difficult to accurately portray the current efficiency of an OR comparative to others, even within similar hospital systems. It is key, therefore, to clearly define and establish variables which are relevant across all systems.

A simple method of calculating OR efficiency is to use overutilization and underutilization times per OR day [11]. Underutilized time is calculated as a percentage of the scheduled time (i.e., one hour underutilized in an 8-hour day = 12.5%) and overutilized time is calculated at

double that rate due to monetary and morale costs associated with unexpected overtime (i.e., one hour overutilized in an 8-hour day = 25%). However, this method only gives a value for inefficiency in the OR without accounting for key measurables which contribute to creating efficiency and which could potentially be improved.

One proposed scoring system, established by Macario [13], included eight (8) metrics for measuring OR efficiency in which each scored from 0-2 points, with higher scores indicating more efficient performance. The eight metrics, along with the highest ranking requirement, are: (1) excess staffing costs (less than 5%); (2) mean start-time tardiness per OR per day (less than 45 min); (3) case cancellation rate (less than 5%); (4) percent of workday with at least one PACU admission delay (less than 10%); (5) mean contribution margin per OR hour (more than $2,000); (6) mean turnover times (less than 25 min); (7) prediction bias per 8 hours of OR time (less than 5 min); (8) percentage of turnovers greater than 60 minutes (less than 10%). Measurables relating to turnover times, first case on-time starts, and case duration estimates were also noted as key performance indicators by Kurtz [14] and Foster [15]. Foster also used other surgery specific measurables including subsequent case on-time starts, preadmission screening for anesthesia, patient-in-to-incision and patient-close-to-out times, surgical checklist compliance, and prime time utilization. Alternatively, the other variables noted by Kurtz included percentage of locations used and times they are used, surgical complications, value-based purchasing, consistency of service, and surgical outcomes—all of which could be influenced by factors outside of surgery, such as administrative decisions, patient factors, or circumstances beyond control. It is preferable, therefore, to refrain from using such measurables to ensure that the outside factors will not influence end results, an outlook echoed by Divatia & Ranganathan [12] about using such outcomes to assess OR efficiency.

2.2. Methods to Improve OR Efficiency

With the key variables in place for measuring OR efficiency, the next step is to find methods of improvement. There have been many reports completed regarding methods of improving OR workings, with many different approaches taken. First, two articles which reviewed groups of other studies will be covered, then individual studies.

Fong et al [16] reviewed 39 previous studies and found a natural separation in the methods used based on the size of interventions undertaken: small (single-operative team), medium (floor or group), or large (institution). Based on their review, they further broke down the interventions taken at each level. Small-scale interventions included redesign of surgical workflow, standardization of instruments and supplies, and the implementation of team huddles before and after each case; medium-scale interventions centred around widespread use of checklists, increase in teaming (use of the same surgical team members), data tracking, and improving surgeons' awareness of costs for disposable instruments; large-scale interventions involved changes and improvements in supply chain management, implementation of specialized personnel and services where volume allows, and space redesign and parallel processing. Methods of measuring hospital supply chain management performance and improvement may be reviewed in a literature study completed by Moons et al [17].

Saleh et al [18] took a different approach by assessing and grouping studies based on their relevance to the operation as either preoperative, intraoperative, or postoperative. In this way, the interventions were separated primarily by the teams affected in each stage of operation. For preoperative, the studies assessed used methods of improvement that focused on communication, staffing, changes to patient flow, and generating earlier patient evaluations and documentation. In the grouping of intraoperative improvements, the methods used included reducing variability

by implementing clinical pathways, modifying staffing for short-procedure rooms, and by utilizing surgeon profiles to maximize throughput. The postoperative improvement studies analyzed utilized methods of improvement such as parallel processing with a dedicated PACU room, improving communication, and implementing a 60-minute standard for turnovers.

An article regarding improvements undertaken at University of Alabama at Birmingham University Hospital from 2004 to 2006 to improve staff retention and increase efficiency after transitioning to an expanded replacement hospital provides examples of how a single facility may undertake multiple projects to meet specific problems [19]. To combat issues of mandatory overtime, unpredictable work hours, and poor staff morale there were a number of improvement initiatives completed, including first case on-time starts, turnover time efficiencies, prime time block use, nursing staff issues, central sterile service issues, and patient safety. Through these initiatives, they were able to "retain critical human resources and restore a supportive environment for the patients, the doctors, and the staff."

Other actions taken to improve OR efficiency coincide with some of the measures mentioned previously, namely turnover time, on-time starts, and scheduling. The reduction of anesthesia turnover time from 65 to 52 minutes was completed for the Clinic for Gynaecology at the University Hospital, Zurich, Switzerland through increased staffing to allow for parallel processing (beginning induction of the next case before the prior one had ended) [20]. At a Veterans Affairs Medical Center in Salt Lake City, Utah, the rate of on-time surgical starts was improved by implementing pre-OR timeouts with a safety checklist alongside a modest performance pay incentive for attending surgeons, resulting in an improvement from a rate of 15% of cases starting on-time to over 72% [21]. Finally, the method taken by the University of Louisville School of Medicine was to improve scheduling of surgeries, as well as preadmission

testing, which had a positive impact on first case on-time starts, OR time utilization, and case volume during peak OR time [22].

These are only a few of the methods that may be taken to improve OR efficiency. In such a complex process as the operating room and associated areas, the opportunities for improvement are endless. In the end, regardless of the method taken the end goal is to provide an improvement on an OR efficiency measurable, such as those described in the previous section.

2.3. Data Driven Improvements and Systems Engineering

In many hospitals, including St. Boniface Hospital, planning and managerial practices are completed by a human planner even though advanced techniques and algorithms from the field of operations research could provide better results [23]. Indeed, when it comes to tackling issues that arise within a system as complex as the OR, the strategy for directing improvements is often determined by "common sense". However, this can be an issue precisely due to the complex nature of the system. Without a properly guided analytical and mathematical approach, it is possible that key issues may be missed or misdiagnosed. In addition, it is possible that an environment of passing blame may arise because different parties involved in the system will have bias regarding what is causing issues.

Although improvements may still be seen without the use of data-driven methods, they may not be the most effective or efficient. One problem faced with implementing these advanced analytical methods for decision making is that although many OR managers are skilled leaders with a healthcare background, they are unlikely to possess the necessary skillset to utilize such mathematical methods [24]. For this reason, it is highly beneficial to have an engineer or similarly mathematically oriented professional as a part of the OR team.

Traditionally, engineers have been involved in healthcare in the form of clinical engineers. In this role, the engineers are responsible primarily for maintenance, training, operation, and safety of instruments, design of procedures or components as required, and optimization of information handling and analysis (though an information technology professional or business may be used for information handling and associated services) [25] [26]. More recently the roles of engineers have shifted to include industrial and systems engineering approaches, though this application of engineering knowledge is not as widespread in healthcare services. From the 2001 report completed by the Committee on the Quality of Health Care in America [27]:

> *"Other world-class businesses, notably those that have received the prestigious Malcolm Baldrige National Quality Award, have embraced many of the tenets of quality improvement described by Deming, Juran, and others, which include the need to improve constantly the system of production and services. Yet few health care organizations have developed successful models of production that reliably deliver basic effective services, much less today's increasingly advanced and complex technologies."*

The lack of reliable production and improvement models designed specifically for healthcare services appears to be due to the inherent complexity and variability that comes from systems steeped in human factors. Indeed, there is some skepticism that methods that were developed for manufacturing and industrial setting could be applicable in such a different setting. However, the application of data-driven and engineering-based improvement methods can be successfully applied in healthcare so long as it reaches a deeper level than simply individual projects; it must become a part of the healthcare culture or else these methods will have only as much success as any other methods that came before and were eventually discarded [28].

In addition, the benefits of using data-driven approaches for tackling problems with efficiency outweigh the challenges to implementing these new methods. From a report on an initiative taken by the UC Davis Medical Center in 2010 to improve OR capacity and patient/staff satisfaction, it was determined that the optimization project resulted in more than simply improvements to efficiency, it "transformed a culture" [29], which—as addressed previously—is key to sustaining changes. The additional benefits that were seen from this shift in culture included strengthened personal relationships, reinforced teamwork and communication, investment of OR staff in process and quality improvement, and a change from blaming others for problems to looking for root causes using data and metrics, which further encouraged staff to feel open to making suggestions and calling for improvements where they were most needed.

Even without completely changing the culture in healthcare, improvements may be seen through the use of data. By posting the percentage of daily on-time starts publicly, LeAnn Northam of the Riley Hospital for Children Indianapolis brought attention to the issue and generated discussion within the various associated departments (anesthesia, nursing, surgeons), which resulted in marked improvements (21% to 60% high, and average of 47.25%) [30].

Data-driven approaches are also useful for reducing and planning for variability. Currently, most OR systems operate on a reactive basis to problems—for example, rescheduling or cancelling surgeries as changes occur and scheduling staff to cover rooms that will run overtime in the middle of the day. This is largely due to an inability of the system to adequately plan for variability in advance. By applying variability theory to account for both common-cause variability (artificial or controlled, i.e., variation in elective surgeries and scheduled cases) and special-cause variability (natural or uncontrolled, i.e., emergency case or unexpected surgical

complications), improvements can be made to all planning and scheduling endeavors. A project undertaken by the Mayo Clinic Florida to apply operational management and variability theory produced significantly positive results in the form of increased throughput without added expense, reduced overtime, improved staff satisfaction, decreased same day changes to the schedule, and overall improved financial performance [31].

2.4. Lean Six Sigma in Healthcare: Outcomes and Challenges

As explained in Section 1.3, Lean Six Sigma is a continuous improvement methodology which has origins in manufacturing, and which utilizes data-driven methods. A focus on improving quality in healthcare at a national level began in the 1990s [32] [33] [34] [35], around the same time that Lean and Six Sigma were developed and began gaining traction in manufacturing. As business began to apply Six Sigma techniques to their customer service sectors—alongside their original manufacturing applications—it became clear that such a technique for improvement is highly applicable to the service industry sectors. Thus, the implementation of manufacturing improvement methods in healthcare was considered.

Some of the earliest articles regarding Six Sigma and Lean in healthcare were in the 1998 article, “Is Health Care Ready for Six Sigma Quality?” [36] and in the 2001 profile piece, “Physician Strives to Create Lean, Clean Health Care Machine” [37]. A compelling article by Chip Caldwell [38] and the associated book [39] from 2005 provide a compelling case for the use of Lean Six Sigma in healthcare and direction for upper management on its application.

Although these methods are now more widely used in healthcare and have many successful applications worldwide, there are still many healthcare systems that stand to learn from these methodologies. As well, though some hospitals may apply Lean Six Sigma to individual projects, the cultural shift which provides the sustainability for such improvements is still to

come. Even though there have been great successes with Lean Six Sigma in healthcare, there are also challenges which are faced and which provide pushback against changing to such a system of improvement.

Projects using Lean, Six Sigma, or a combination of both are plentiful in the literature. A strength of Lean Six Sigma is that it can be applied to any and all elements of the OR; this also makes it more difficult to review because the literature and case studies available are extremely varied. Thus, the focus here will be on outcomes assessed based on primary Lean Six Sigma project articles and from summary information for reviews of literature currently available, on information about hospitals and health services which have achieved a complete cultural shift toward quality and continuous improvement, and the key challenges faced in implementing these improvement methodologies.

2.4.1. Primary Studies

From research published in the last fifteen years, 32 cases which utilized Lean, Six Sigma, or Lean Six Sigma in an OR environment were reviewed for the outcomes associated with the project implementation. In all cases, significant improvements were realized. A breakdown of the outcomes which were seen in multiple studies is given in Table 2-1 (outcomes seen in two or less cases were not included). Reduction in patient length of stay was the most prevalent improvement measured, with improvements to costs, turnover times, wait times, and first case on-time starts as outcomes for improvement also appearing in five or more studies. Three to four cases measured improvement to infection rate, throughput, process flow, staff morale, and turnaround time. Each outcome was then separated into one of the following categories: health & safety, quality & satisfaction, process efficiency, and cost. Overall, it seems that all OR

improvements could be assigned as one of these categories, although some measures could fall into multiple (i.e., reduction in overtime could be associated with cost or quality & satisfaction).

Table 2-1 Recorded Outcomes of Lean Six Sigma in Healthcare

Outcome	Reference	Count	Category	Description
length of stay	[40] [41] [42] [43] [44] [45] [46] [47] [48] [49]	10	health & safety	Decrease in patient length of stay
cost	[41] [44] [45] [47] [50] [51] [52]	7	cost	Decrease in operational costs
turnover time	[49] [50] [53] [54] [55] [56]	6	process efficiency	Decrease the time between last patient out and next patient in
wait times	[51] [57] [58] [59] [60] [61]	6	quality & satisfaction	Decrease wait-times for patients to receive care
first case on-time start	[53] [62] [63] [64] [65]	5	process efficiency	Improve rate at which first surgical cases start on-time
infection rate	[45] [66] [67] [68]	4	health & safety	Decrease the risk of infections post-surgery
throughput	[59] [60] [69] [70]	4	process efficiency	Increase in the number of cases which can be accommodated per room
process flow	[48] [51] [64] [71]	4	process efficiency	Rework the process to improve flow
morale	[50] [51] [56]	3	quality & satisfaction	Improve staff morale and satisfaction
turnaround time	[50] [55] [61]	3	process efficiency	Decrease time between surgical dressing end and surgical incision for the subsequent patient

2.4.2. Reviews of Studies

An article by Deblois and Lepanto provided a methodological review of reviews [72]. After searching articles published between 1999 and 2015, seven reviews published between 2009 and 2012 were included in their assessment. From their breakdown, four key effects were repeated throughout the Lean Six Sigma project reviews: health outcomes, processes, quality, and costs. This categorization of outcomes is comparable to those given to the review of primary studies above, and also to those discussed in the review completed by D'Andreamatteo et al [73], which

reviewed 109 empirical studies and separated the impacts as primarily relating to safety, productivity, clinical quality, and cost. In both reviews, it was determined that Lean Six Sigma is a promising approach to generating improvements in healthcare, however there are challenges to implementation and more research is needed to further explore the limits of this methodology in healthcare.

2.4.3. Organizational Implementation

Although reviews and individual successes indicate the usefulness of Lean Six Sigma and continuous improvement methodologies in healthcare, the real proof of success comes from places where a cultural shift has occurred and there is universal acceptance and application of the improvement methodology within the organization. Four hospital systems within the United States which have had success at this level are assessed below: Virginia Mason Medical Centre, Mayo Clinic, Cleveland Clinic, and ThedaCare.

The Virginia Mason Medical Centre in Seattle, Washington began their OpEx journey in 2001 with a visit to Japan, after which they began to form their own continuous quality improvement program called the Virginia Mason Production System (VMPS) based on the Toyota Production System/Lean principles (Figure 2-1 shows a simplified version of VMPS in a diagram). They now have hundreds of staff trained in their system, with 20 full-time employees dedicated to implementing VMPS projects and training. One element which allows VMPS to thrive is their "no-layoff rule"—any staff which have their role eliminated by efficiency improvements are guaranteed to be redeployed in another area. They have had many positive results from VMPS projects including decreasing staff walking distances by 38%, reducing parts travel by 77%, reducing inventory and lead time by half, increasing productivity from 44-93% in various areas, reducing patient waits for lab results by 85%, and improvements in cost in the

amounts of $500,000 overtime savings, $1 million in lowered operating costs, and a cost savings in budgeted capital between $12-15 million [74].

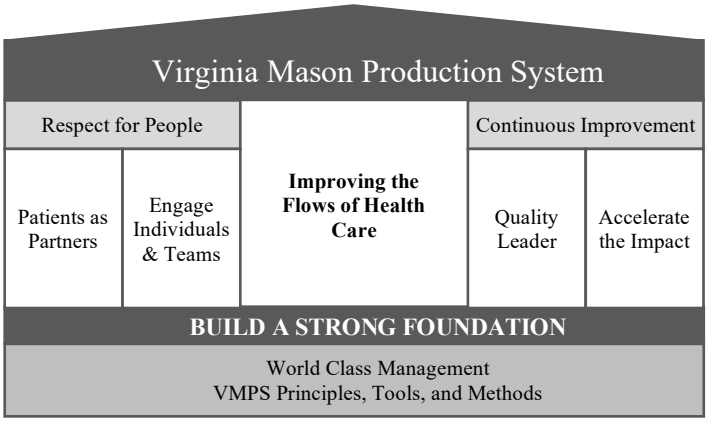


Figure 2-1 Simplified Virginia Mason Production System House [75]

The Mayo Clinic in Rochester, Minnesota also developed their own OpEx system which they call the Mayo Clinic Quality Construct (MCQC). The MCQC consists of four fundamental elements—culture, infrastructure, engineering, and execution—with the overarching mission being to provide the best care possible to their patients [75]. Figure 2-2 gives a simplified diagram of the interactions between the elements of their quality construct. In 2006, the Mayo Clinic Quality Academy was established in order to support the quality improvement efforts. The MCQC incorporates a combination of improvement approaches, including "Just Do It", Plan-Do-Study-Act (PDSA), Lean, and Six Sigma [76]. The quality and efficiency of the Mayo Clinic is evident from holding the title of top hospital in the United States, as ranked in the 2020-2021 U.S. News & World Report. The Cleveland Clinic, in Cleveland, Ohio, holds the second rank in the same list [77] and is also known for their quality improvement efforts.

The Cleveland Clinic Improvement Model (CCIM), similar to the MCQC, uses tools from various improvement methodologies including huddles, visual management, Plan-Do-Check-Act (PDCA), Kaizen, "Just Do It", standardization, and 5S. The CCIM is broken down into four elements: organizational alignment, visual management, problem solving, and standardization

[78]. The application of this improvement model at the Cleveland Clinic is assisted by their continuous improvement department, which is made up of professionals from Lean, Six Sigma, and project management backgrounds [79].

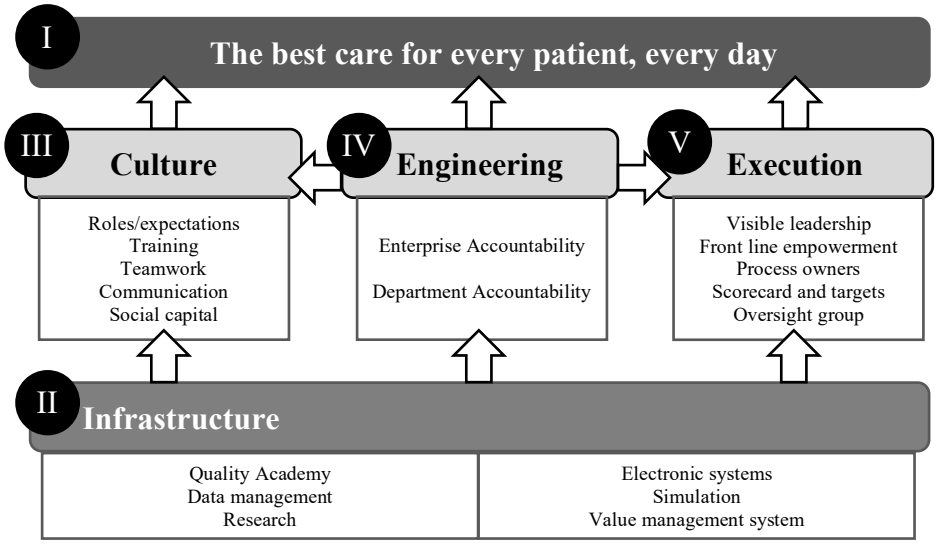


Figure 2-2 Simplified Mayo Clinic Quality Construct [76]

The final organizational application of Lean to be assessed is ThedaCare, which is a Wisconsin health system comprised of seven hospitals and numerous clinics and services. Their Lean journey began in 2003, at which time they developed the ThedaCare Improvement System (TIS) which utilize value stream mapping (VSM), rapid improvement events (RIE), and process projects. Similar to Virginia Mason's system, ThedaCare has a "no-layoff" rule incorporated in the TIS [80]. While the TIS produced improvements—they achieved bottom-line savings of $25 million by 2009—they did not reach all the goals which had been set. It was determined that the main issue was a lack of change in management to accompany the changes that had been implemented with TIS. This led to the creation of the ThedaCare Business Performance System, a system of management implemented to reduce variability among managers, generate daily improvements, and prioritize improvement projects [81].

The thing that each of these systems has in common is that they were able to combine aspects of different improvement systems and personalize it to the needs of their organization. This personalization, along with the involvement of front-line workers and all levels of management is what has helped each of these organizations to be successful in their continuous quality improvement initiatives.

2.4.4. Challenges with Lean Six Sigma Implementation

Regardless of the potential, healthcare holds unique barriers in successful implementation of Lean Six Sigma and other continuous improvement initiatives. Firstly, Lean and Six Sigma were created for a manufacturing environment, so there is often pushback against these methodologies due to the belief that healthcare is too different from manufacturing for the same methods to be applicable to both. Indeed, it is easier to apply such techniques to manufacturing due to the linear nature of product development and healthcare does present challenges in the way of being highly complex and dependent on human variables. Therefore, the first challenge faced in implementing Lean Six Sigma is in the perception of those being asked to adopt the new method. At Virginia Mason, some physicians and management personnel even quit when the new VMPS was implemented. Others resisted the initial implementation but were swayed as the gains became evident; this caused some of the initial nay-sayers to become vocal advocates of the system after time [82]. Resistance to change—both in management and on the front lines—is a common barrier due in part to perception, as discussed, but also due to fear of job loss [53] [54] [65] [72].

Another issue is the training and terminology needed to apply Lean Six Sigma. There are many definitions which have Japanese origins and require additional training to understand. Much as a layperson would not readily understand the medical terminology of *sphenopalatine ganglioneuralgia* (brain freeze), neither would a medical professional understand terms such as

muda (waste), *gemba* (workplace), *kaizen* (improvement), DPMO (defects per million opportunities), or *kanban* (signboard). As well, referring to the patient as both the "customer" and "product", and viewing a hospital as a business can be concepts which are not readily accepted in the healthcare setting. Thus, the training of staff, both in terminologies and methodologies, is another key challenge of implementation which has been documented [43] [72] [76].

Other challenges assessed in the literature include lack of resources [53] [82], lack of experts in healthcare [71], failure to commit to the change and involve all levels of the organization [71] [76] [83], issues with healthcare hierarchal and organizational design [71] [83], and problems due to the complexity of the system and differing goals between departments [71] [76] [83].

2.5. Lean at St. Boniface Hospital

In 2008, Lean was introduced to St. Boniface Hospital and the Transformation Team was created with Dr. Michel Tétreault, former St. Boniface President and CEO, as its champion. Mentors with Lean experience from Standard Aero, ThedaCare, and Simpler Consulting aided the Transformation Team at the start of their journey. The first actions of the Transformation Team were hiring initiatives and creating value streams for ER acute coronary syndrome [84] and acute care surgical service, after which two additional front-line staff were added and more VSM and RIE were completed in areas such as resuscitation room organization [85], surgical flow, medicine flow, and supply chain. Through these projects, positive results were achieved in improving ER results, direct admittance to acute care surgical service, checklists and documentation, and fall prevention, among other improvements.

Following the initial momentum of these improvements, came resistance from the front-line staff to changes and a loss of some of the improvements applied. One of the reasons for losses in

the early improvements was due to the fact that, after the Transformation Team withdrew from areas where improvements were made following completion of the projects, the staff remaining were not trained to sustain and continue the improvements. From 2012 to 2015, there was a change in the Transformation Team in terms of stabilization and shifting their focus toward training and developing skills in St. Boniface Hospital staff, rather than in direct participation in transformation projects. In 2015, ThedaCare consulting centre Catalysis aided with defining the management system for Lean at St. Boniface and ensuring that the Transformation Team could give the front line the ability to solve problems and send information higher up as needed. Since that time, the focus has remained on training of staff and also on generating continuous improvements of the systems and processes involved in pilot areas of pharmacy and cardiac surgery. As of 2019, approximately 60% of staff have participated in training initiatives and the Transformation Team continues to evolve and grow, with new staff and directorship.

Next steps for St. Boniface Hospital in its Lean journey include continuing to promote engagement at all staff levels in improvement projects, and training and creating a lasting culture of improvement throughout the hospital. This is the most difficult stage of the journey as it requires a level of acceptance and commitment from all areas that has not been required before.

2.6. Chapter Summary

Within this chapter, many different and viable ways of measuring OR efficiency were assessed (Section 2.1). It is important to consider what is most relevant for the project at hand and define the measurables prior to the start of any improvement project. This allows for clarity during the project and also ease in tracking improvements. In addition, remaining consistent between projects can allow for improved communication between teams and better

comparability of results, both within the hospital and when comparing results to outside sources and other studies.

When it comes to methods for improvement, as stated by Saleh et al [18]:

> *"There seems to be no clear single answer to the question of how to improve operating-room processing and throughput. Rather, there are a number of individual strategies directed at improving particular aspects of operating-room processing."*

And indeed, many documented strategies for making improvements in the OR were found and reviewed (Section 2.2). This is why it is particularly important to determine which areas should be addressed first in order to provide the greatest benefit. A summary of OR efficiency measures and methods of improvement reviewed are shown in Table 2-2.

Table 2-2 Operating Room Efficiency Measures and Improvement Methods

Efficiency Measures	**Methods of Improvement**
OR utilization time	Workflow redesign
Staffing costs	Standardization
First case on-time starts	Huddles
Case start-time tardiness	Checklists
Cancellation rate	Teaming
PACU admission delay	Data tracking
Contribution margin	Cost awareness
Turnover time	Supply chain
Prediction bias	Specialized services
Percent long turnovers	Space redesign
Pre-admission screening rate	Parallel processing
Patient in – incision time	Communication
Patient close – out time	Staffing
Surgical checklist compliance	Patient flow
	Patient evaluations
	Reducing variability
	Maximize throughput
	Scheduling

The benefits of a data-driven or engineering approach were also discussed (Section 2.3). Data-driven methods and systems engineering are tools which can be leveraged to greatly

improve efficiency within healthcare and the OR in particular can benefit due to the many complex interactions involved. Data can help to determine true root causes that may otherwise be obscured by bias. Having a team member that is an engineer, or in a similar field, is key to ensuring that such methods can be properly applied. The use of Lean Six Sigma (or other continuous improvement methodologies) is a prime example for how such data-driven techniques can be applied in a systematic and proven manner for success.

Also reviewed were many case studies and articles which cover the successes and challenges in applying Lean Six Sigma in healthcare (Section 2.4). Positive outcomes realized fell under the categories of health & safety, quality & satisfaction, process efficiency, and cost. The true success of continuous quality improvement in healthcare, though, comes from places like Virginia Mason Medical Centre, Mayo Clinic, Cleveland Clinic, and ThedaCare, all of whom have managed to create and cultivate their own culture of improvement within their organization.

Creating a change in healthcare culture is a big challenge, and it is only one of many challenges which were covered. Perception, training, resources, commitment, organizational hierarchy, and complexity are all challenges that are faced when attempting to apply a new continuous improvement methodology to healthcare, particularly one such as Lean Six Sigma which is so strongly descended from manufacturing backgrounds.

The use of Lean and Six Sigma methods within St. Boniface Hospital was the final information covered in this chapter (Section 2.5). Within St. Boniface, the implementation of Lean and Six Sigma methods is in an earlier stage and the hospital has yet to reach the stage of complete cultural immersion. However, great strides are being made with the Transformation Team within the hospital and, hopefully, in time the entirety of the organization will be well-versed and able to participate actively in continuous improvement initiatives.

3. IMPROVEMENT METHOD: LEAN SIX SIGMA PROJECT

The Lean Six Sigma improvement project completed as a part of this thesis took place at St. Boniface Hospital. The improvement area was the OR and all associated processes.

3.1. Define the Initiative—Project Charter

Generating a Project Charter to define the project goals, team, and other elements was the first step taken. The full Project Charter can be found in **Error! Reference source not found.**. It was broken into the following sections: background, business case, project objective, scope/boundaries, schedule/milestones, team/resources, risk, and risk mitigation/action plans.

The project background gave a brief introduction to Lean Six Sigma and identified consistently being on-time at end-of-day in the OR as the end goal. The business case supporting this goal was established in the next section, noting that 51.5% of end-of-day cases ended more that 15 minutes late and resulted in 2152 hours of overtime. Late cases and high levels of overtime result not only in increased costs but also poor patient flow, cancelled cases, issues with resource utilization, and poor patient experience, proving that this is a crucial area for improvement. Thus, the project objective was set at reducing the number of late end-of-day surgeries from 125 per month to an average of 30 per month.

The scope of the project was established next, with elective surgeries, slating procedures, staff roles, and all associated processes within scope and emergency cases, surgical processes, and staff hours being out of scope. The timeline for Lean Six Sigma milestones was also set, though adjustments were made throughout the course of the project to account for delays.

Roles of the project team members were clarified at this time as well. As stated previously in Section 1.3.3.1, the team roles for this project were different from a normal Lean Six Sigma project. A normal Lean Six Sigma project includes a Lean Six Sigma Green or Black Belt

certified leader, team members from the process being improved with Lean Six Sigma training, and an executive team to support the process. This project was completed entirely by the team leader (myself), with leaders in the OR environment acting as the executive team. Therefore, all planning, analysis, and work was completed by the team leader, with the executive team giving approval and input as required as well as directing the team leader to staff members with knowledge of the OR processes.

Four risks were identified at the start of the project and mitigation measures were outlined. The first risk identified was delays due to staff being on vacation, which was mitigated by allocating enough time within each section that such delays were accounted for and requiring alternative contact information be provided if necessary. The second risk identified was falling behind schedule. The mitigation measures in place to correct for this risk were implementing weekly team meetings, increasing communication with the process owners, and allocating tasks to OR/L2PO staff members.

The third risk assessed was that the data consultant was taking a leave of absence. Due to this being a planned leave, an alternate contact was able to be identified and additional time was allotted when requesting data during that period. The final risk identified was that of staff being resistant to change. To ensure acceptance of changes, a plan was made to include affected staff in the problem-solving process, create a compelling argument for all changes, and provide support and coaching to staff if necessary. A report detailing all findings and data would also be made available, providing additional justification for any changes implemented.

After creating the Project Charter, it was reviewed and approved by the executive team. Each revision of the Project Charter was also approved by the executive team, aside from the final revision which could not be approved due to complications caused by COVID-19.

3.2. Characterize the Process

With the OR efficiency under analysis and reducing overtime as the improvement goal, all OR processes and associated processes required assessment to improve understanding of the improvement team. To achieve this on-site observation, interviews, and process mapping were completed. Data from April 2017 to April 2018 were also analyzed after being filtered and edited, the steps of which are detailed in APPENDIX C. Images of the original spaghetti diagram maps and associated interview notes, and sticky-note mapping are included in APPENDIX D.

3.2.1. Process Characterization and Linkages

The first step taken in this section involved understanding how the processes and steps within the hospital and OR fit together. To accomplish this, first the process tiers were identified, as shown in Figure 3-1, to give an overview of the system for the purposes of this project.

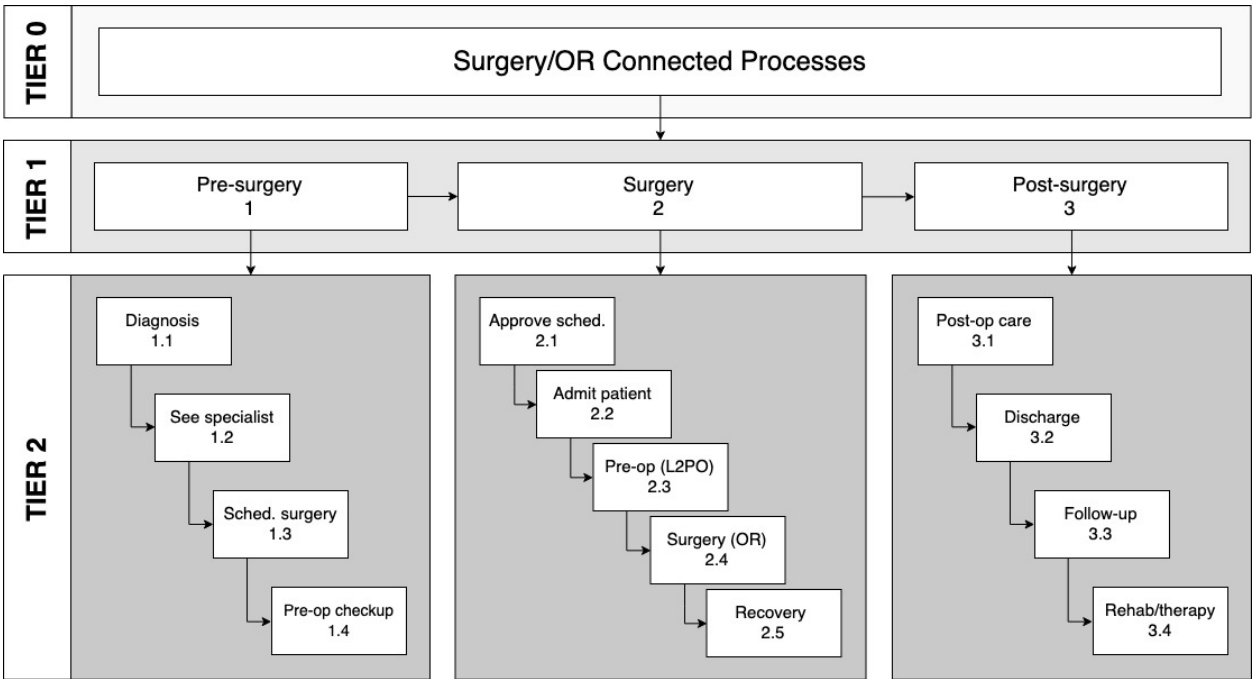


Figure 3-1 OR Process Tiers

Surgery (the second step of Tier 1) was identified as the principal area for improvement, primarily within steps 2.1-2.4. These are the areas directly related to the operating rooms on the

day of operation: scheduling, admitting, pre-op, and surgery. Recovery, aside from the transportation of the patient from the OR into recovery, was not more closely assessed, as the project goal of reducing overtime and late end-of-day surgeries is more highly influenced by pre-operative and operation areas. Slating (a.k.a. surgery scheduling) was also assessed as it impacts the flow of surgeries.

Next, the different clinical programs, supporting departments, and administrative departments were considered for how they work together to keep the specified surgical areas running (Figure 3-2). These connections were used to aide in defining how the process works.

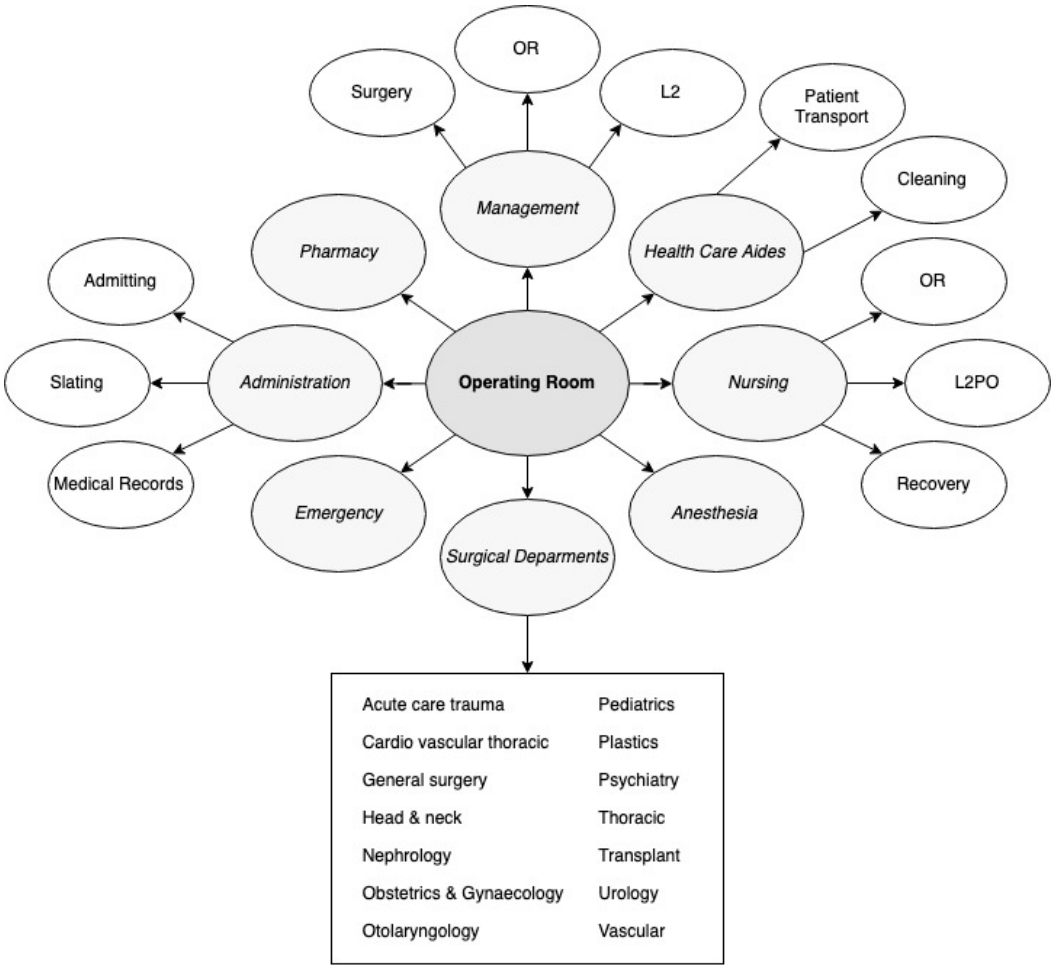


Figure 3-2 Connections between hospital departments to OR

3.2.2. Process Workings and Performance Issues

With the tiers and focus areas identified, a SIPOC Analysis was completed, beginning with customer Critical to Quality characteristics (CTQs), which were identified in CTQ trees (Figure 3-3) to clarify the customer requirements in measurable terms. Three primary requirements were chosen with relation to OR overtime and efficiency: information and paperwork, resources, and surgery.

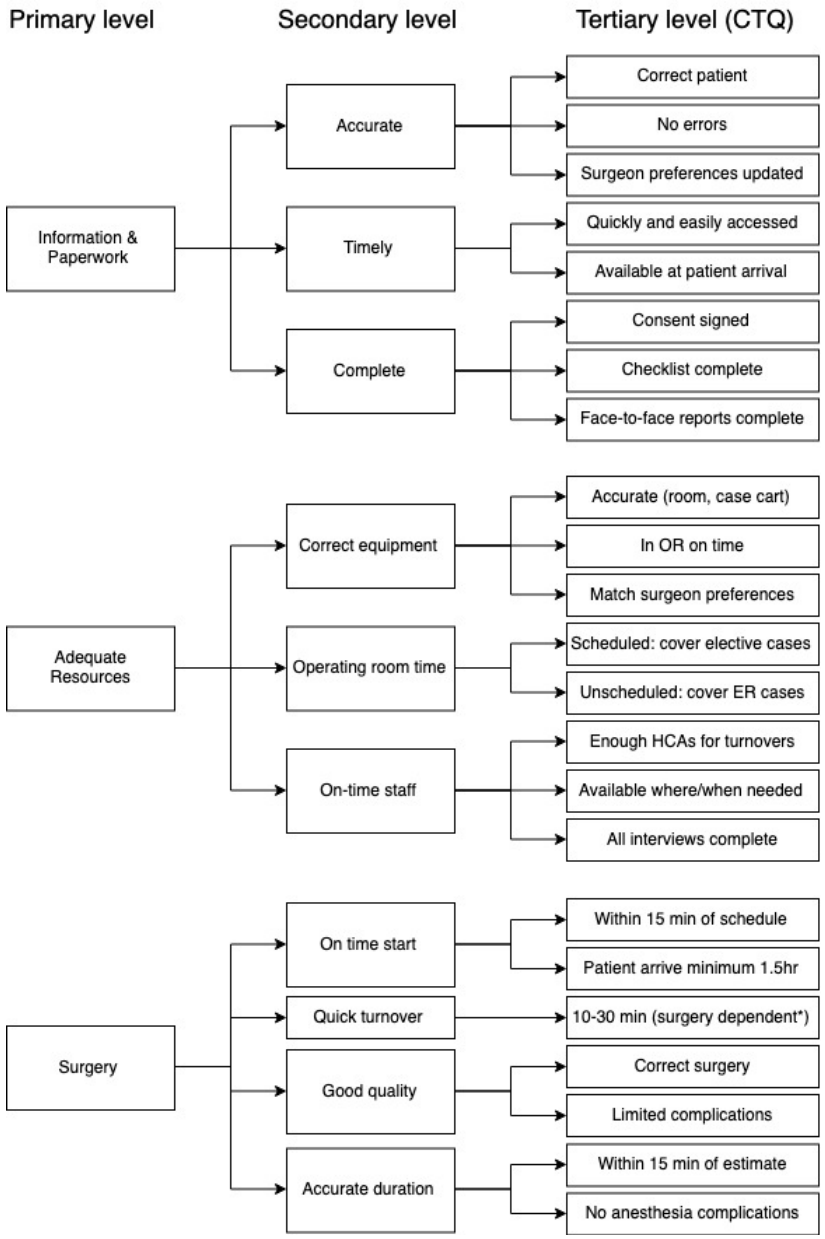


Figure 3-3 CTQ Trees

The CTQs with time oriented measurables were identified as key lagging measures of OR efficiency, including cancellations, overtime, start times, case duration estimates, and turnovers. The defects per unit (DPU) associated with each measurable were compiled in Table 3-1. This covers the process function with the product or service given and the associated CTQ, defect, unit, and DPU values. This gives a current view of errors within the OR processes.

Table 3-1 Defects Per Unit (DPU) Matrix

Function	Product/ Service	Customer Expectations	Defect	Unit	DPU
Surgical Department	Management/ Finance	No OR overtime	Overtime (> 0 min)	Each room (for each day)	1522 defects 2515 units = 0.605 DPU
L2PO/OR	Surgery	First case on time start	Late (> 0 min)	First elective cases	1793 defects 2501 units = 0.717 DPU
L2PO/OR	Surgery	Elective case on time start	Late (> 0 min)	Completed elective case	4269 defects 6235 units = 0.685 DPU
OR	Surgery	Accurate daily duration	Over/under time (> 15 min)	Each room (for each day)	2042 defects 2515 units = 0.812 DPU
Turnover	Clean-up	Timely	Extended time (>25 min)	Primetime room turnover	1279 defects 3711 units = 0.345 DPU
OR	Surgery	Completed	Case cancelled	Scheduled elective case	1004 defects 7239 units = 0.139 DPU

The SIPOC Analysis was compiled to bring together all aspects of the process understanding that were undertaken, shown in Figure 3-4. The lagging measures and customer needs for the SIPOC Analysis were pulled from the previous DPU Matrix. The customer expectations from the defect matrix are the performance issues to be pursued for improvement.

From the SIPOC flowchart, an integrated flowchart (Figure 3-5) was created to develop a more detailed view of the process steps. This visualizes the many tasks that the staff must complete within the OR to contribute to the process success. A specific completion time for each task could not be included as they may vary depending on the patient, procedure, or workload for number of staff available.

Figure 3-4 SIPOC Analysis

Suppliers	Input	Process	Outputs	Customers
Client (Patients) Administrators Slating clerks Nurses HCA Surgeons Anesthesiologist	Patient (pre-surgery) Patient admittance Full slate Patient/Room prep Turnover, Supplies, Aid Surgery, Interviews Anesthesia, Interviews	Name: Operating Rooms Purpose: perform surgical procedures Owners: Lance Barber Dawn Affleck Lorena Thiesson Tamara Miller	Patient (post-surgery) Waste materials Room cleaned (ready for next patient) Actual surgery duration/times	Client (Patients) Family of patients Recovery

Process Steps					Result (Lagging) Measures	Customer Needs
Admit patient - patient arrival - checkin	→ **Pre-op (L2PO)** - medication/IV - pre-op checklist - interviews	→ **Prep surgery** - prep room - OR briefing - prep anesthesia	→ **Perform surgery** - anesthesia - surgery - wrap-up	→ **Recovery/Turnover** - moved to recovery - clean room - clean anesthesia	Overtime rooms (#, time) Cancelled cases	**Patient Needs:** Successful surgery On-time start/finish Info/communication **Staff Needs:** Resources available Full staffing Accurate schedule Accurate surgeon pref.

Process (Leading) Measures	Patient arrival time # of patients	Cycle time # of surgeries Time of day	Cycle time % equip in place % case cart correct % pref sheets correct	Anesthesia type Surgery type/duration # of secondary procedures	# of HCA	**Priorities for Improvement**
Sources of Variation	Patient transportation Family/friends	Medications Patient requirements Consent/forms Communication	Equipment available Surgeon/anes. available Surgeon preferences Communication	Complications Schedule inaccurate	Bed availability	First case on-time start Turnover times Case duration accuracy Surgeon preferences Staff timeliness Communication/Information
Impact	Late arrival Delay going into L2PO Cancellation	Delay going into OR	Delay starting surgery	Delay ending surgery Overtime/undertime Cancellation	Delay into next case	

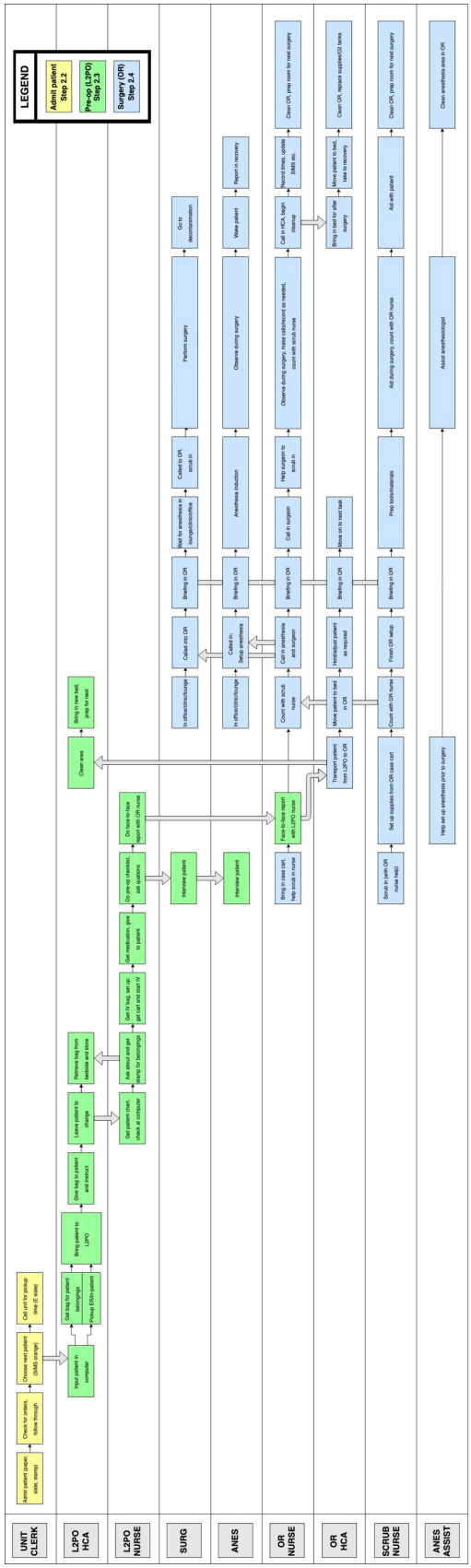


Figure 3-5 OR Integrated Flowchart

3.3. Improve

This improvement project focused on improving OR efficiency so that overtime work—identified as a CTQ characteristic and the goal for improvement—would be reduced. The current level of rooms running overtime is unacceptable, with over 60% of rooms running elective surgeries going overtime (of those, 85% ran overtime by greater than 15 minutes). There are many current process problems needing improvement to close this gap, shown in Figure 3-6. These areas were assessed for impact and prioritization (Table 3-2) as potential projects, with approval from the executive team.

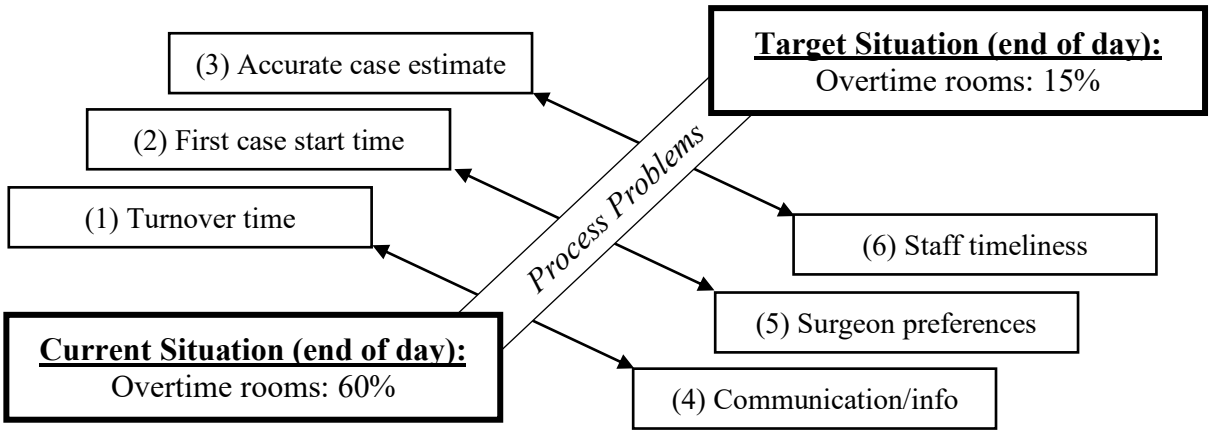


Figure 3-6 Performance Gap

Table 3-2 Prioritization Matrix

OBJECTIVES		**POTENTIAL PROJECTS (see Figure 3-6)**					
Objective	**Weight**	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
On-time end of day	0.4	7	9	9	3	5	9
Patient satisfaction	0.1	3	7	5	5	1	7
Staff satisfaction	0.1	5	5	7	9	9	5
Increase case throughput	0.1	9	3	7	3	3	5
Improve quality	0.1	1	3	3	9	3	1
Improve safety	0.1	1	3	3	9	1	1
Data collection	0.1	7	1	3	1	3	1
Weighted average (benefits)		5.4	5.8	6.4	4.8	4.0	5.6
Expected timeline		2 years	2 years	2 years	1 year	6 months	1 year

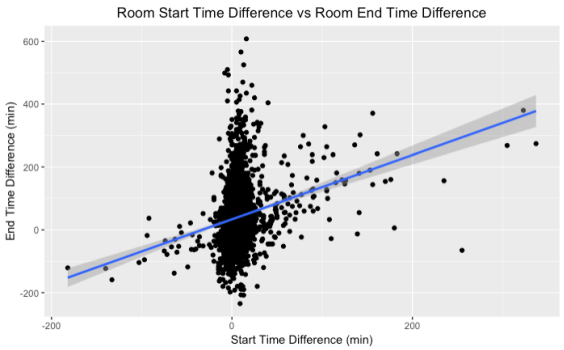
*Note: cost was not considered as a factor as the work was being completed on a volunteer basis as part of a master's thesis

To calculate the weighted average, the relationships between the potential project and hospital objectives were assessed on a scale from 0 to 9, with 0 indicating no relationship and 9 indicating a strong relationship. Of the objectives, having an on-time end of day was weighted heavily while the other objectives were evenly weighted.

At the start of the project, first case start times were recommended by the executive team as the focus for improvement. However, to properly apply the Lean Six Sigma methodology all other process areas also needed to be considered for improvement, resulting in the six potential projects given above. Of the potential projects, two were prime for improvement: first case on-time starts and accurate case duration estimation. Both had high DPU values (Table 3-1), high prioritization based on the weighted averages (Table 3-2), and high interest within the hospital for improving these measures. After assessing these two areas further, it was determined that the accurate case estimation would be pursued, for two reasons.

Firstly, improvement in case estimates would have a higher impact on reducing overtime. This conclusion was based on the knowledge that inaccurate case duration estimates impact cases throughout the day and could negatively impact other areas of interest such as turnover times and staff timeliness; with cases not ending on time, staff may have difficulty adequately preparing for turnovers or starting on the next cases.

Secondly, it was determined that first case start times do not have as large an impact on end of day lateness as expected. The correlation of room start time and end time error was assessed using a scatter plot and Pearson correlation test (Figure 3-7). The correlation coefficient is 0.251 with a p-value of $2.2e^{-16}$, thus start time does have a slight positive correlation to end time error. However, it is also clear based on the spread of the points in the scatter plot off of the linear band that the majority of rooms do not have a direct linear correlation between start and end times.




```
        Pearson's product-moment correlation

data:  testdata$startDiff and testdata$endDiff
t = 12.966, df = 2506, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.2136897 0.2870558
sample estimates:
      cor
0.2507328
```

Figure 3-7 Correlation between room start and end difference from schedule

In addition, as shown in Table 3-3, 49% of rooms which begin late also end late, however, on average the rooms with a late start will end 39 minutes later than they started (shown in Figure 3-8), thus the late beginning only accounts for only 26% of the total lateness for these rooms. As well, over half of the rooms that start on-time will also finish late, by an average of 14 minutes. With on-time rooms starting an average of 6 minutes early, this is a 20-minute delay which accrues over the course of the day due to other factors. Overall, while the first case start time does impact the room end time, the case slating and estimates of case duration were deemed to have more room for improvement and greater impact on overtime reduction than improving first case start times.

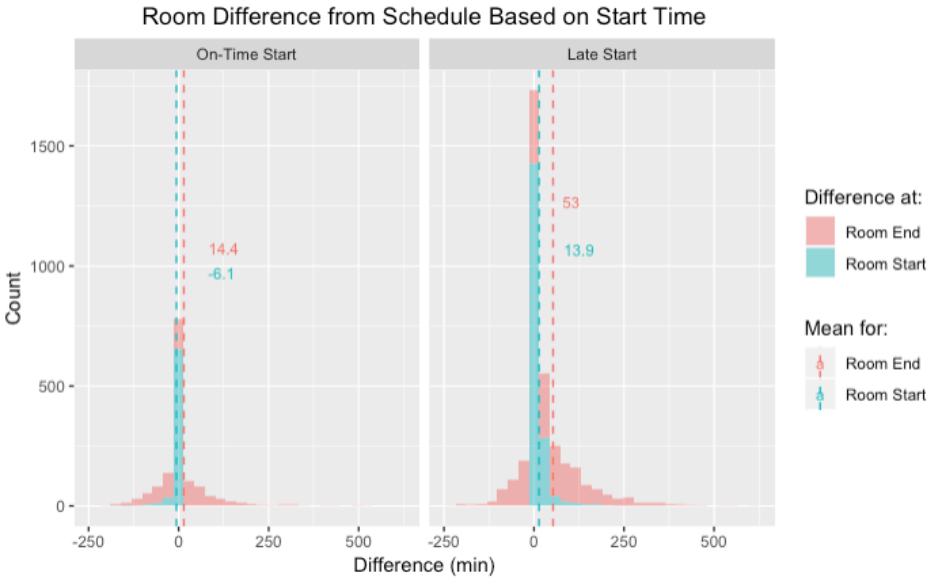


Figure 3-8 Rooms start and end differences from schedule grouped by on-time/late start

Table 3-3 On-Time vs Late Room Start and End times

	LATE START	ON-TIME START	TOTAL
LATE END	1240 (49.4%)	375 (15.0%)	1615
ON TIME END	560 (22.3%)	333 (13.3%)	893
TOTAL	1800	708	2508

Improving turnover time was also considered as a focus. However, it was deemed to be a future endeavour as current methods of record keeping did not directly note turnover times, nor were causes for extended turnovers consistently documented. Instead, turnover times were extrapolated from "patient out" to "next patient in" time values. This calculated value could be affected by many different processes, as can be seen by the many different roles and steps which encompass the task of moving from one patient to another, shown in the integrated flow chart (Figure 3-5) previously. With so many varied possibilities for error, additional data tracking is needed to determine which turnover tasks would most benefit from improvement.

3.3.1. Define (Establish the Focus)

From the previous section, it was determined that improving the accuracy of case and room duration estimates would be the focus for the improvement project. In order to reduce OR overtime, the overall daily room durations and case durations must be properly estimated.

3.3.1.1. Describe and Verify the Performance Gap

Currently, the daily room duration estimates are at unacceptable levels as only 18.8% of rooms finish within 15 minutes of the expected end time. Both over- and under-utilization of the OR is an error that should be improved upon, though only over-utilization counts towards the over-time hours accrued by the OR. For individual cases, 32.8% of case durations were within 15 minutes of the estimated durations. For this calculation, the average turnover time of 28

minutes was assumed and added to the case duration as the scheduled time includes turnovers while the actual case durations did not.

The Short-Term Process Sigma was calculated using defects per million opportunities (DPMO) in Table 3-4. These values were calculated for the room duration error as well as for the case duration error as a baseline for performance. A shift of 1.5 was assumed for calculating the short-term capability due to the use of long-term data. An ideal process would be within 6 Sigma, meaning only 3.4 DPMO. The current process is far out of this range, at 0.61 Sigma and 1.05 Sigma for room duration error and case duration error, respectively.

Table 3-4 Process Sigma Data and Calculations

Unit	Room	Case
Defect	Error > ±15 min	Error > ±15 min
Defect opportunity [O]	1	1
Units [N]	2515	6235
Errors [D]	2042	4192
Yield [(N*O – D)/(N*O)]	18.8%	32.8%
DPMO [(1-Yield)*1,000,000]	811,928	672,334
Short term process sigma (1.5 shift)	0.61	1.05

3.3.1.2. Develop the Project Objective and Improvement Plan

Moving forward, the objective is to improve the accuracy of case duration estimates used to create the daily room slates in order to allow the rooms to end on-schedule, thus reducing the amount of underutilized OR time and reducing the amount of required overtime. In addition, the improved case estimates may allow for more accurate preparation of staff and patients through the course of the day. The plan is to assess the current process used to create the daily room slates (Section 3.3.2: Measure) and determine what root causes are affecting the case estimate accuracy (Section 3.3.3: Analyze). Action will then be taken to improve the case estimates (Section 3.3.4: Improve).

As there are many surgical departments and procedures included within the OR, it was determined that the project scope would be narrowed to surgeries which occur frequently enough to give statistically significant values and have a high likelihood of occurring during the project timeline to allow for new data to be tested following improvement. This allows for a more in-depth look at the data and processes involved in specific cases and surgical departments. The process used for choosing the surgeries can be found in detail in APPENDIX E. Three factors were considered when choosing cases: 1) frequency of procedure, 2) room for improvement, and 3) surgical departments and surgeons affected. The six procedures chosen are from cardiac (coronary artery bypass graft (CABG), mitral valve replacement/repair (MVR), aortic valve replacement/repair (AVR)), vascular (carotid endarterectomy (END)), and obstetrics-gynecology/OB-GYN (total abdominal hysterectomy (HAT), diagnostic hysteroscopy (HYSD)). The baseline process sigma was calculated to reflect the starting point for each of these six procedures (Table 3-5). The average turnover time for each procedure was added to each case duration to account for the turnover time that is included within the scheduled value and not in the actual case durations.

Table 3-5 Process Sigma Data and Calculations for Chosen Procedures

Unit	CABG	AVR	MVR	END	HAT	HYSD
Defect	Error > ±15 min					
Defect opportunity [O]	1					
Units [N]	477	164	82	73	159	182
Errors [D]	417	140	73	53	126	81
Yield [(N*O – D)/(N*O)*100]	12.6%	14.6%	11.0%	27.4%	20.8%	55.5%
DPMO [DPO*1,000,000]	874,214	853,659	890,244	726,027	792,453	445,055
Short term process sigma (1.5 shift)	0.35	0.45	0.27	0.90	0.69	1.64

After action has been taken, the progress will be measured by assessing improvements to the baseline values. Due to the limited timeline and resources, improvements will not be

standardized and implemented throughout the OR. However, based on the tested level of improvement, recommendations will be made for future actions and conclusions, and reflection on the success and challenges of the project will be completed (Section 3.3.5: Control).

3.3.2. Measure (Examine the Current Situation)

In this section, the current process for slating and generating case estimates is assessed through the use of a process map and baseline data analysis, relevant CTQs are defined, and a failure mode and effects analysis is used to assess ways in which the estimates could fail. A cause-and-effects matrix was deemed unnecessary at this time due to there being only a single input and output requirement (input: surgeon estimate; output requirement: estimate accuracy). A capability analysis was also not completed due to the high likelihood of variability due to special causes (i.e., surgeon thoughts/decisions, surgical complications) impacting outcomes.

The process used to slate surgeries varies between cardiac cases and other elective cases. All elective cases (minus cardiac, C-sections, and electroconvulsive therapy (ECT)) are scheduled using PART, which is an electronic version of paper booking cards. Cardiac cases are scheduled using slate management and are often not given a duration; the cardiac rooms have a ten-hour day and one-to-two cases to complete, so the surgeries are simply assigned a time slot by the scheduler, commonly at 5 hours per case. For all slated cases, certain information must be included: visit number, medical record number, patient demographic, procedure, surgeon, and patient status (day surgery, same day admit, inpatient admit).

The operating rooms and dates are scheduled based on surgeon and service department. If a timeslot is not filled it will be given to another surgeon or department, so it is important for the surgeons to fill the timeslots allotted to them. The steps of the slating process for elective and cardiac cases are shown in Figure 3-9.

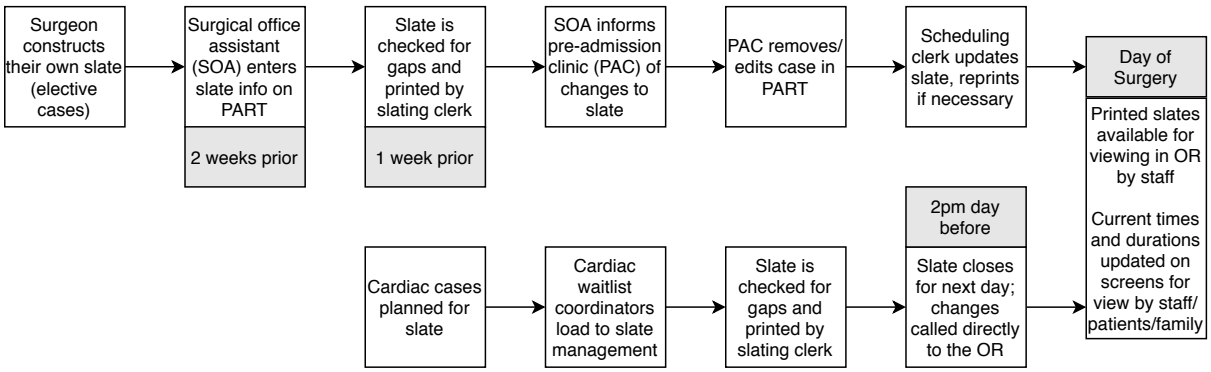


Figure 3-9 Slating process map, elective and cardiac streams

The CTQ associated with this project which requires an operational definition is "accurate duration estimate". In the current process, the duration estimate includes both the case duration, which is the time that the patient enters the OR to the time they exit the OR, and turnover time, which is the time between one patient exiting and the next entering the OR. To measure the duration accuracy, the estimate is compared to data recorded during surgery. If the actual duration is within 15 minutes of the estimate it is considered accurate, as initially shown in Figure 3-3. It does not have to be on-time to have an accurate duration. An overtime duration occurs when the case runs more than 15 minutes past the estimate, while an undertime duration occurs when the case finishes more than 15 minutes early. Due to the estimates containing both case duration and turnover time, when comparing the duration estimate with actual case duration, turnover time must be either removed from the estimate or added to the actual duration.

Baseline data for each of the chosen surgeries were initially assessed during their selection using $\bar{X}$-R control charts, histograms, and boxplots (available in APPENDIX E). Initial values for each of the six surgeries are given in Table 3-6, both without and with the extrapolated turnover times added. The three cardiac surgeries have a high positive mean error between their scheduled and actual durations indicating that they often run overtime, and they also have large standard deviations. On the other hand, the single vascular surgery and two OB-GYN surgeries

have negative average error between the scheduled and actual durations, indicating that they often end earlier than scheduled. By choosing surgeries with such varied baselines and issues, we will be able to observe how well the solutions proposed work for surgeries of different types.

Table 3-6 Baseline values of selected procedures without and (with turnovers)

Surgical Dept.	Primary Procedure Description	Cases	Mean Error (min)	SD of Error (min)	Max. Error (min)	Min. Error (min)	Sum of Error (min)
Cardiac	Bypass graft, coronary artery (CABG)	477	66.0 (76.3)	91.0 (87.2)	438 (438)	-110 (-98)	31477 (36413)
Cardiac	Replacement/repair valve, mitral (MVR)	82	86.0 (94.7)	118.2 (116.8)	458 (458)	-212 (-212)	7056 (7772)
Cardiac	Replacement/repair valve, aortic (AVR)	164	35.9 (45.1)	95.3 (92.7)	395 (395)	-101 (-85)	5889 (7402)
Vascular	Endarterectomy, carotid (END)	73	-40.6 (-25.2)	37.7 (44.3)	54 (75)	-117 (-117)	-2967 (-1843)
OB-GYN	Hysterectomy, abdominal total (HAT)	159	-4.9 (16.6)	50.1 (49.2)	264 (264)	-127 (-107)	-785 (2645)
OB-GYN	Hysteroscopy (diagnostic) (HYSD)	182	-24.4 (-7.2)	17.1 (27.7)	17 (162)	-70 (-70)	-4432 (-1312)

There are only two ways for the surgery duration estimate to fail: by being too long or too short. These modes of failure are assessed in the FMEA (Table 3-7). Although there are many possible causes for the case duration being incorrect (complications, delays, tardiness of staff, procedure cut short), an inaccurate method for estimating durations is a possible cause for both types of failure, thus is a prime place to start in assessing and improving function. In addition, there is not currently a process in place to control for error in duration estimates—the estimates submitted by the surgeon are simply entered into the slate as given. Building in a control process could also aide in reducing errors in case duration estimates. Although short and long estimates are both sources of error, controlling for short estimates (resulting in overtime) is of more importance than preventing long estimates (resulting in undertime), because the effects and costs of overtime are greater than those of undertime.

Table 3-7 Failure Mode and Effects Analysis

Process Step	Potential Failure Mode	Potential Failure Effects	SEV *	Potential Causes	OCC **	Current Process Controls	DET ***	RPN ****	Recommended Action
Surgery Duration Estimate	Too short	Overtime Late next case Cancelled cases Patient/Family worry Next patient anxious Hard to plan turnover Staff planning issues No time for breaks	7	<u>Common Causes:</u> Inaccurate method Wait for material Error in prep Long turnover time Staff tardiness <u>Special Causes:</u> Complications Delays in other areas	9	No controls in place for surgery estimates; slating clerk enters times as provided by surgeons (controls in other areas to prevent errors/delays day of surgery, such as with case-carts/ room prep)	9	567	Short estimates have greater impact than long, but both share a common cause: inaccurate methods. Changing the process to create more accurate estimates could improve both. In addition, a process control system which can assess estimates for error likelihood (i.e., compare to mean, standard deviation, 75th percentile) could help prevent error.
	Too long	Undertime Next patient not ready Room underutilized	4	<u>Common Causes:</u> Inaccurate method <u>Special Causes:</u> Procedure cut short Easier procedure	5		9	180	

*SEV = severity of the failure effects, ranked from 1 to 10; high severity ranking indicates high risk

**OCC = occurrence of the failure, ranked from 1 to 10; high occurrence ranking indicates high potential to fail

***DET = detection capability for the failure mode, ranked from 1 to 10; high detection ranking indicates low capability for detection

****RPN = SEV*OCC*DET = risk priority number, assessment of risk priority level with values ranging from 1 to 1000; high RPN indicates high priority

Therefore, the objective is to improve the accuracy of case duration estimates used to create the daily room slates in order to allow the rooms to end on-schedule. There are a number of possible strategies for improvement that may be considered to accomplish this goal, shown in Table 3-8.

Table 3-8 Strategies and Measures

#	Strategy	Measure
1	*Breakdown estimates into each part:*	
	Improve/include separate turnover time estimate	Turnover time accuracy
	Improve/include separate anesthesia time estimate	Anesthesia time accuracy
	Improve/include procedure duration estimate	Procedure duration accuracy
	Improve/include wrap-up time estimate	Wrap-up time accuracy
2	*Breakdown estimates into surgery and turnover:*	
	Improve/include separate turnover time estimate	Turnover time accuracy
	Improve case duration estimate (patient in to out)	Case duration accuracy
3	*Improve total surgery estimate (patient in to next in)*	Case duration + Turnover time accuracy
4	*Improve daily room duration estimate*	First case start – Last case end duration accuracy

3.3.3. Analyze (Analyze the Causes)

In this section, the current process is analyzed to determine the root cause of the problems so that action may be taken. For this project, this involved the use of a cause-and-effect (AKA fishbone) diagram to brainstorm the causes behind errors for case duration estimates.

Due to the high variability of patient conditions and surgery elements, it is expected that some error will be present. However, minimizing the amount of error for individual cases and creating a slate in which the error over- and under-time balances out is key to ensuring rooms will end on time, with the highest degree of effective OR usage. Based on information gathered from staff involved in scheduling surgeries, a cause-and-effect diagram was created (Figure 3-10) with the causes for error in case duration split into six categories: slating methods, surgeon, surgery, materials/equipment, patient, and day-of causes.

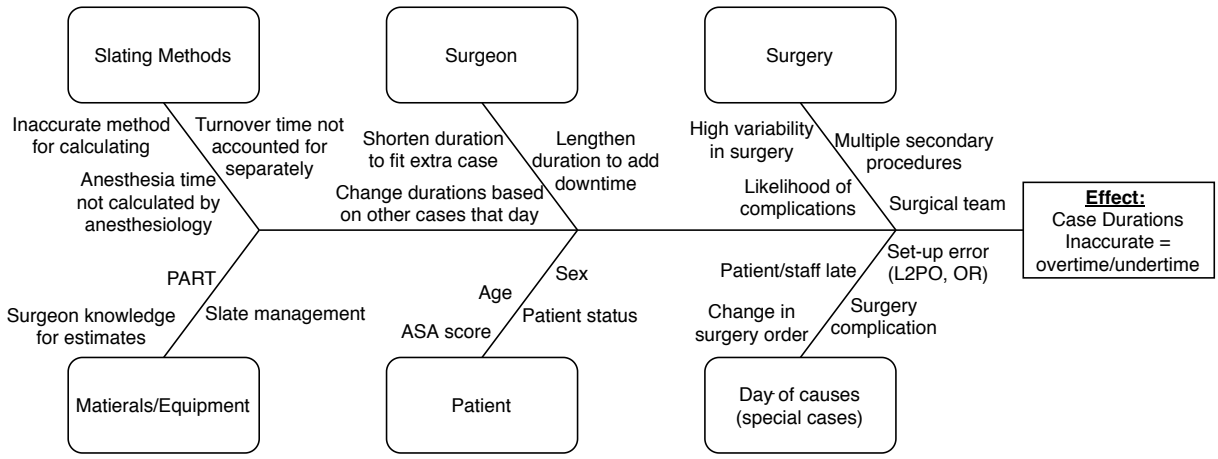


Figure 3-10 Cause-and-Effect (AKA Fishbone) Diagram

The day-of causes increase the actual duration past what would be expected, such as with late staff or patients, surgery complications, set-up errors, and changes to the slate. These would also all be classified as special cases/causes, meaning that their occurrence should be controlled and minimized when possible but that they are not the leading cause of consistent error. These could be looked into during future projects to further improve the efficiency of the OR.

The surgery and patient categories both include variables which impact the surgery duration but cannot be changed, only used to aid in creating the estimate. More secondary procedures, high likelihood of complications, high American Society of Anesthesia (ASA) score, or older patients are all likely to increase surgery duration whereas as a highly skilled surgical team which often works together, low risk surgeries, and patients with a low ASA score may result in a shorter duration. Some of these variables are interrelated; the ASA score is based on patient age, health, and other patient factors while the likelihood of complications is dependent on the type of surgery, number of procedures, surgical team, and patient health. As the current method for creating the estimates rests solely with the surgeons responsible for the surgery, it is up to each surgeon how they choose to use this information to generate their estimates.

The causes under the categories of surgeon and materials/equipment also relate to the slating method category because the current slating method relies on surgeon knowledge and choices—to purposely over/under-book durations or otherwise adjust case duration estimates—which are then entered into either PART or slate management. As such, the root cause behind the high rate of error in duration estimates is due to the current methods in place for slating. Relying only on the surgeon's estimate is a flawed system. Surgeons may be unaware of data trends for past cases, as well as values for turnover and anesthesia times. They also may adjust their schedules with different goals in place than scheduling accuracy: increasing number of patients seen for surgery, having extra time between cases, or adding time to one surgery to provide a buffer in case another runs late. This reasoning is known only by the surgeon and cannot be anticipated by other staff, therefore, by building it into the slate it can negatively affect efficiency on a case-by-case basis even if the total room duration is accurately estimated.

To test the current slating method, scatter plots depicting the scheduled durations against the actual durations were assessed (Figure 3-11) along with a count of how many cases ran on-time, over-time, and under-time for each procedure (Table 3-9) both without and with turnover times included. Bands ±15 minutes from the scheduled duration are included in the scatter plots to show the on-time window. Based on the analysis, the current slating method is inaccurate; with actual turnover times included, the most accurate procedure was HYSD with 48% of cases finishing on-time and cardiac cases ranking as the least accurate at 12.8%, 8.5%, and 12.8% of cases finishing on-time for CAGB, MVR, and AVR respectively.

A new slating method which can take into account issues such as surgeon bias, utilize information on patient and case variability, and provide more accurate estimates per case and thus per room is needed.

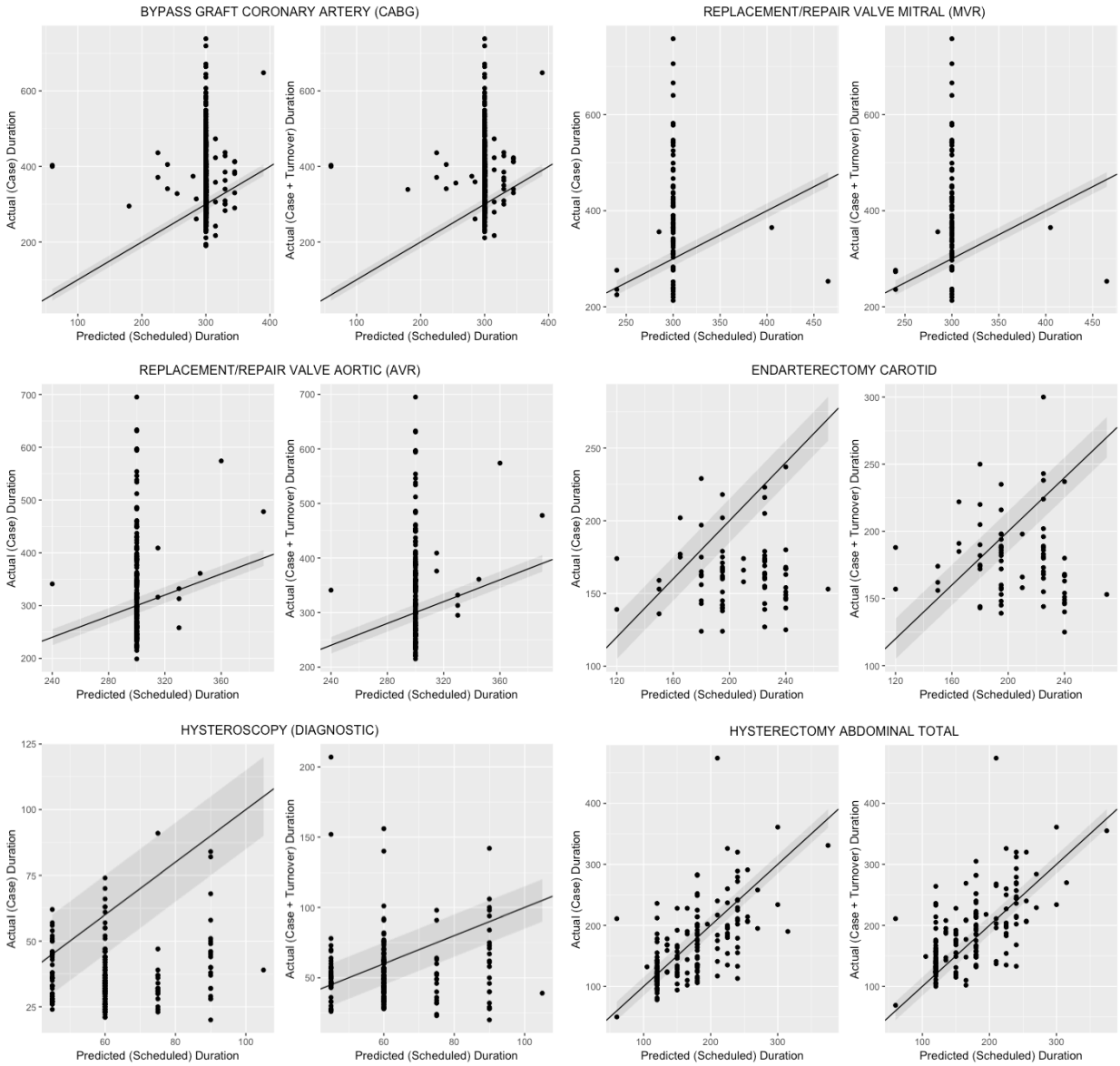


Figure 3-11 Scatter plots with 15-minute bands

Table 3-9 Count of cases which were on-time, overtime, and undertime

		Case vs Prediction			Case + Turnover vs Prediction		
Primary Procedure Description	**Cases**	**On-time**	**Over**	**Under**	**Ontime**	**Over**	**Under**
Bypass graft coronary artery	477	68	322	87	61	352	64
Replacement/repair valve mitral	82	8	58	16	7	61	14
Replacement/repair valve aortic	164	24	80	60	21	92	51
Endarterectomy carotid	73	11	6	56	20	13	40
Hysteroscopy (diagnostic)	182	46	2	134	88	22	72
Hysterectomy abdominal total	159	50	44	65	43	77	49

3.3.4. Improve (Act on the Causes, Study the Results)

This section of the improvement process is broken into two parts: acting on the causes (on a small scale) and studying the results.

3.3.4.1. Act on the Causes

Improvements to the slating method are needed to rectify the current error rate of the case duration estimates. For this project, the focus is on improving estimate accuracy for the six surgeries identified, which are across three surgical departments. Actions which were determined to be options for improve slating accuracy are given in Table 3-10, along with the advantages and disadvantages associated with each. The two actions of changing from surgeon estimates to a data-driven slating method using either historical averages or predictive modeling were ranked as the top choices for action. Updating surgeons with historical data for their most common procedures was ranked as the second choice and adding separate duration estimates from anesthesiology staff and health care aides (HCAs) for anesthesia and turnovers, respectively, was ranked as the last choice. These rankings were approved by the executive team.

Table 3-10 Possible actions

Action	Advantages	Disadvantages	Rank
Provide updates to surgeons of data for their most common surgeries to improve their estimate accuracy	• Minimal changes • Easy to implement • No cost	• Still relying on human estimate • No anesthesia/turnover • Surgeons may disregard • Still no process control	2
Add separate estimates for anesthesia, procedure, and turnover from anesthesiologist, surgeon, and HCAs	• Includes anesthesia/turnover • Doesn't impact surgeons • Each area responsible	• Even more human estimates • More info from more sources • Work to combine all info • Still no process control	3
Switch to a data-driven method of estimate generation using historical averages	• Uses past data to ↑ accuracy • One system to make estimate • Eliminate human error • Built in process control	• Surgeons lose control • Big change • Wouldn't include all factors	1
Switch to a data-driven method of estimate generation using predictive modeling	• Uses past data to ↑ accuracy • Includes many factors • Eliminate human error • Built in process control	• Surgeons lose control • Big change • May need multiple models	1

Adding estimates from anesthesiology and HCAs may sound reasonable to ensure that times for anesthesia and turnovers are included in the estimate, but it would make the process more complicated as multiple groups would need to be consulted to create the slate and it increases the opportunity for human error and bias to be introduced, which is why this is the least desirable action. Changing to a data-driven method is the best choice because it is a method which is not susceptible to human error or bias, as the surgeon-based estimates are. However, it is a more complex solution and may require more input from the hospital before it can be implemented properly.

Giving surgeons updates on—or easy access to—their historical data is an easy solution, but still relies on the surgeon to make the estimate. Providing surgeons updates on their historical data does not require any additional changes to the system aside from the compilation and distribution of data. This is a suitable solution for the interim while testing is done for data-driven methods. After discussion, it was decided that this short-term solution would be implemented outside of the project scope (details provided in APPENDIX F).

For this project, the action to be taken is to implement data-driven methods for case duration estimates and assess improvements to accuracy. The plan is to create new models for slating estimates and test them against the current method of surgeon estimates.

Difficulties with generating accurate surgical case duration estimates is not a new problem and many methods and studies on improving estimates have been documented. Some of the methods that have been tested and compared include both traditional (surgeon estimates, historical data) and modern approaches (predictive modeling, machine learning). Variables to be considered for improving accuracy include the various parts of surgery (preparation, anesthesia, procedure, wrap-up, turnover), patient health factors, surgical factors, and surgeon estimates and

insights. Additional details in regard to research on surgical duration models and variables are included in Appendix G.1, along with more information on the three types of predictive methods chosen for use in this project: simple moving average (SMA), linear mixed effects (LME), and random forest (RF). The following sections of APPENDIX G document the data editing (Appendix G.2) and exploration (Appendix G.3) completed prior to creating the models, steps taken to generate the models (Appendix G.4), results of testing the models (Appendix G.5), and R code and functions created for the models (Appendix G.6).

The SMA model was built using the historical data of the previous 19 cases of the same procedure. The only variable used to separate the cases for prediction was the primary procedure. The average of the previous 19 case durations was used to generate the next case duration estimate. Therefore, once 19 cases had been completed an estimate could be created for all future occurrences.

Two LME models were created: LME-case and LME-part. Approximately 80% of the available data were used to create these models. LME-case used a single LME model to predict the case duration, while LME-part was a compilation of two LME models (anesthesia and procedure duration) and two constant values (preparation and wrap-up duration). The LME models utilized all 13 available variables and log-transformed duration data, and the constant values were calculated as the average duration for each primary procedure.

The random forest model also utilized all 13 variables to generate predictions and was created using the same 80% of cases as the LME models. However, due to the type of RF model used, a maximum of 32 factor levels was allowed for each variable, therefore three of the variables (surgeon, anesthesiologist, anesthesia type) had to be altered such that least common factors were grouped together. This was also done for the variable “rooms” to account for the

very infrequent use of certain rooms. A summary of the details for each of the four models is provided in Table 3-11. With the models created, the next step is to generate predictions, compare them to the current method, and study the results.

Table 3-11 Model details

	SMA	**LME-case**	**LME-parts**	**Random Forest**
Cases used to create model/estimate	19 / 1071	861 / 1071	861 / 1071	861 / 1071
Outputs	Case duration	Case duration	Preparation duration Anesthesia duration Procedure duration Wrap-up duration	Case duration
Duration data transform	None	Log	PREP/WRAP: Log ANES/PROC: None	None
Number of variables	1	13	PREP/WRAP: 1 ANES/PROC: 13	13
Variable information	Primary procedure: • CABG • AVR • MVR • END • HAT • HYSD	Fixed: Scheduled duration Num. of procedures Day of the week ASA Score Sex Patient status Age Random: Case service Room Anesthesia Anesthesia type Nested: (case service) Primary procedure Surgeon	PREP: Constant value, primary procedure ANES: Same as LME-case PROC: Same as LME-case WRAP: Constant value, primary procedure	Constants: Scheduled duration Num. of procedures Age Categorical: (factors) Primary procedure (6) Case service (3) Day of the week (5) ASA Score (5) Sex (2) Patient status (3) Altered: (factors) Room (7) Surgeon (29) Anesthesiologist (29) Anesthesia type (17)

3.3.4.2. Study the Results

Two rounds of testing were completed, the first which used the 20% of data remaining after creating the models (after filtering out the first 19 cases of each procedure which had no SMA predictions) and the second which used new data from April 2018 to July 2019. A total of 210 cases were used in the first round of testing and 1314 in the second round. The values used to compare the estimation methods are percent of cases which finished on-time, percent of cases

improved from the surgeon estimate, process sigma value, error distribution values (mean, standard deviation, minimum, maximum), and error sums (overtime, undertime, sum of absolute error, sum of error). All values can be found in Appendix G.5.

The primary goals were to maximize the number of on-time cases and reduce overtime (or undertime for END and HYSD procedures, which had a greater issue with undertime cases). For most procedures where overtime was reduced by the use of a model instead of the surgeon estimate, the undertime value increased and vice-versa. This is to be expected and is acceptable so long as the model does not cause extreme error in the opposite direction.

A low sum of error indicates that the error is evenly distributed (an equal amount of over- and under-time). This is especially useful for procedures which are not the only case of the day, as an even distribution of error is more likely to cancel out over the course of several surgeries. An equal distribution of error is also preferrable to having extremely large overtime error, as overtime is can be costly, stressful for staff, and result in cancelled cases.

When comparing predictions for all of the procedures combined, the four new models all performed better than the surgeon estimates. Overtime minutes were reduced by 50% or more from surgeon estimates, and the rise in undertime minutes was less than the overtime saved—only 14-50% of the overtime error was added back as undertime error. In addition, the error distribution drastically improved using the models, with even the model which showed the least improvement resulting in a 65% improvement in the sum of error. The rate of cases which finished on-time also improved 27-63% from surgeon estimated values. When comparing the procedures separately, some variation was seen in performance of the models versus surgeon estimates. Table 3-12 shows which method of prediction performed best and worst in the areas of on-time cases, over/under-time minutes, and sum of error.

Table 3-12 Model results, best and worst

		Models – Initial Data		*Models – New data*	
Procedure	**Value**	**Best**	**Worst**	**Best**	**Worst**
All cases	On-time cases	LME-case	Surgeon	LME-case	Surgeon
	Overtime error	RF	Surgeon	SMA	Surgeon
	Sum of error	RF	Surgeon	SMA	Surgeon
CABG	On-time cases	RF	Surgeon	RF	Surgeon
	Overtime error	RF	Surgeon	SMA	Surgeon
	Sum of error	RF	Surgeon	SMA	Surgeon
AVR	On-time cases	Surgeon	RF	LME-part, RF	Surgeon
	Overtime error	RF	Surgeon	SMA	Surgeon
	Sum of error	RF	Surgeon	SMA	Surgeon
MVR	On-time cases	(All models)	Surgeon	RF	LME-part
	Overtime error	LME-part	Surgeon	LME-case	Surgeon
	Sum of error	LME-part	Surgeon	LME-case	Surgeon
END	On-time cases	(All but RF)	RF	RF	Surgeon
	Undertime error	LME-case	Surgeon	LME-part	Surgeon
	Sum of error	RF	Surgeon	RF	Surgeon
HAT	On-time cases	LME-case	SMA, RF	RF	SMA
	Overtime error	RF	SMA	RF	Surgeon
	Sum of error	Surgeon	RF	SMA	Surgeon
HYSD	On-time cases	RF	Surgeon	LME-case	Surgeon
	Undertime error	LME-case	Surgeon	LME-part	Surgeon
	Sum of error	RF	Surgeon	SMA	Surgeon

Based on the results of the two rounds of testing, analysis of the model performance was completed for each procedure. However, due to some of the procedures having few cases for comparison in the first round of testing, the interpretations and findings discussed here for each procedure will be based only on the second round of testing.

The CABG procedure showed a drastic improvement in both on-time cases and overtime reduction. On-time cases increased 225-270% from the surgeon estimates (though only 5% of cases were on-time with surgeon estimates to begin with). The SMA model showed the lowest sum of error and the greatest reduction in overtime by almost 70% from surgeon estimates, but also the greatest increase in undertime, which was equal to 42% of the total overtime minutes

saved and 16.5 times the amount of surgeon estimated undertime. The LME and RF models—though they showed only 50-55% reduction in overtime minutes—had a far lower increase in undertime, equal to only 15% of the overtime minutes saved or 5 times the surgeon estimated amount. Although this resulted in the LME-case, LME-part, and RF models having a larger sum of error due to greater overtime values, this may be preferrable for this procedure as it is often the only case of the day—or only one of two—and therefore the equal distribution of error is of less importance (this also applies to the other two cardiac procedures, AVR and MVR).

The AVR procedure showed an improvement for all models with 15-58% increase in on-time cases. As with the CABG procedure, the SMA model showed the greatest reduction in overtime error (52% reduction from surgeon estimate), but the greatest increase in undertime error (5 times the surgeon error). Alternatively, LME and RF models resulted in only slightly less overtime improvement (43-48%) and a significantly lower increase in undertime (1.7-2 times the surgeon error).

For the MVR procedure, only the SMA and RF models produced an improvement in on-time cases over the surgeon estimate (50% and 75%, respectively). The LME-case and LME-part models actually had 13% and 38% fewer on-time cases than the surgeon estimate. However, all models showed an improvement over the surgeon estimate for overtime error and sum of error, with the two LME models performing best overall with a 66-68% decrease in overtime (better than the SMA model) and with very low sums of error.

The only vascular procedure included in this study, END, had all models show an improvement over the surgeon estimates. As previously stated, this procedure had a greater issue with undertime error than with overtime. The LME and RF models improved the number of on-time cases by 70% , and SMA by 60%. The undertime error was improved between 55-72%

across the models, with the SMA and RF models also showing a 28% reduction in overtime and the LME models showing only a slight (<5%) increase in overtime. The surgeon sum of error for this procedure was 1108 minutes of undertime, which the SMA and RF models were able to reduce by 83-90%. The two LME models also reduced the sum of error in minutes by over half, however, the sum of error for these models were in overtime minutes, which is more costly than undertime and thus not an ideal outcome.

For the HAT procedure, only three of the models outperformed the surgeon estimates. A 10-15% improvement of on-time cases compared to the surgeon estimate was seen for the LME and RF models. The SMA model did not improve on the surgeon estimates for this procedure as it had an 18% decrease in on-time cases and the only reason it showed an improvement in sum of error was because it produced only a very slight (<5%) reduction in overtime error while more than doubling the undertime error such that it was nearly equal to the overtime error. Although the LME and RF models also increased the undertime error by nearly double, this was less than that added by the SMA model, and the LME and RF models were able to reduce overtime error by 42-51% from the surgeon estimates.

The HYSD procedure, as with the END procedure, had an issue with undertime error rather than overtime. This procedure began with the highest rate of on-time cases (59% with surgeon estimate) and the models were able to improve it by 38-55%, such that over 90% of cases finished on time for the best performing models in this area (LME-case and LME-part). This shows that an accurate surgeon estimate can greatly benefit the function of a predictive model so that an even more accurate estimate can be reached. All the models reduced the undertime error for this procedure. The LME models reduced undertime error the most (75-81% improvement) with only a slight (6%) increase in overtime. Although the RF model showed the most

improvement in reducing overtime error, it was less successful than the LME models at reducing undertime error (improved only 36%) which was the primary issue for this procedure. The SMA showed the same issue as with previous procedures in that the sum of error was very low due to over- and under-time errors being nearly equal, not by greatly reducing error (though it was still an improvement over surgeon estimates).

For all models, the process sigma was calculated. During the first stage of the improvement process the short-term process sigma was calculated as a baseline measure (Table 3-5) based on surgeon estimates from April 2018-April 2019. Normally, comparing to the baseline is necessary after implementing a change in order to assess improvements in the process's ability, because the old system is no longer in use. However, for this project the models were not implemented and the current process was still in use thus comparing the performance of the models to the surgeon estimates for the same data is more telling because they have the same source, rather than comparing it to the initial baseline data. Both original baseline sigma and new surgeon estimate sigma values are compared with the sigma values for the four models in Table 3-13, with the models which performed better than both the baseline and current surgeon estimates shaded green, the best performance shaded a darker green, no shading given to values which were better than the current surgeon estimates but not the baseline, and the values which were worse than both the current and baseline surgeon values shaded red.

Table 3-13 Short-Term Process Sigma (1.5 Shift) Data: Baseline and After Improvement

Model	CABG	AVR	MVR	END	HAT	HYSD	Data
Baseline surgeon estimate	0.35	0.45	0.27	0.90	0.69	1.64	Apr17-Apr18
Current surgeon estimate	-0.14	0.22	0.07	0.81	0.88	1.73	Apr18-Jul19
Simple moving average	0.52	0.30	0.30	1.22	0.73	2.47	Apr18-Jul19
Linear mixed effect, case	0.58	0.44	0.00	1.28	0.99	2.89	Apr18-Jul19
Linear mixed effect, part	0.57	0.50	-0.16	1.28	0.96	2.79	Apr18-Jul19
Random forest	0.59	0.50	0.40	1.28	1.00	2.41	Apr18-Jul19

For the surgeon estimates, the process sigma decreased during the second stage of testing for the three cardiac cases (CABG, AVR, MVR) and the vascular case (END), while it improved slightly for the two OB-GYN cases (HAT, HYSD). The new models performed better than the surgeon estimates in nearly all instances, with the exception of the LME-case and LME-part models for the MVR procedure, which had lower values than both the baseline and current surgeon estimate, and the SMA and LME-case models for the AVR procedure, which had lower process sigma values than the baseline but better than the current surgeon estimate values.

Overall, it is clear that the models are more accurate than the surgeon estimates for all six procedures. This is especially clear when looking at the estimate versus actual duration scatter plots for the procedures in Figure 3-12 (from Figure F-13 in the appendices) as the models more closely follow and are more evenly distributed about the slope (slope = 1). When it comes to determining which model to use there is some difficulty because no single model is the best in every area, so the choice depends on the type of improvement wanted and the format of the model itself. In this case, I recommend using the LME-case model or the SMA model.

The LME-case model is simpler than LME-part, thus would require less editing for errors when adding new cases to the model to improve its performance, and it is not limited by the factors used in training as the RF model is. This limiting of factors could be especially burdensome for the RF model as surgical staff change but is not an issue with the LME model. In addition, the LME-case model performs better than the SMA model in most areas. Although the SMA model often has a lower sum of error, when looking at the scatter plots (Figure 3-12) it is clear that the cause for this is not more accurate predictions but a more even distribution around the slope. Even so, the SMA model would still be a decent replacement for the surgeon estimates because it is extremely easy to implement and could be easily applied to all procedures.

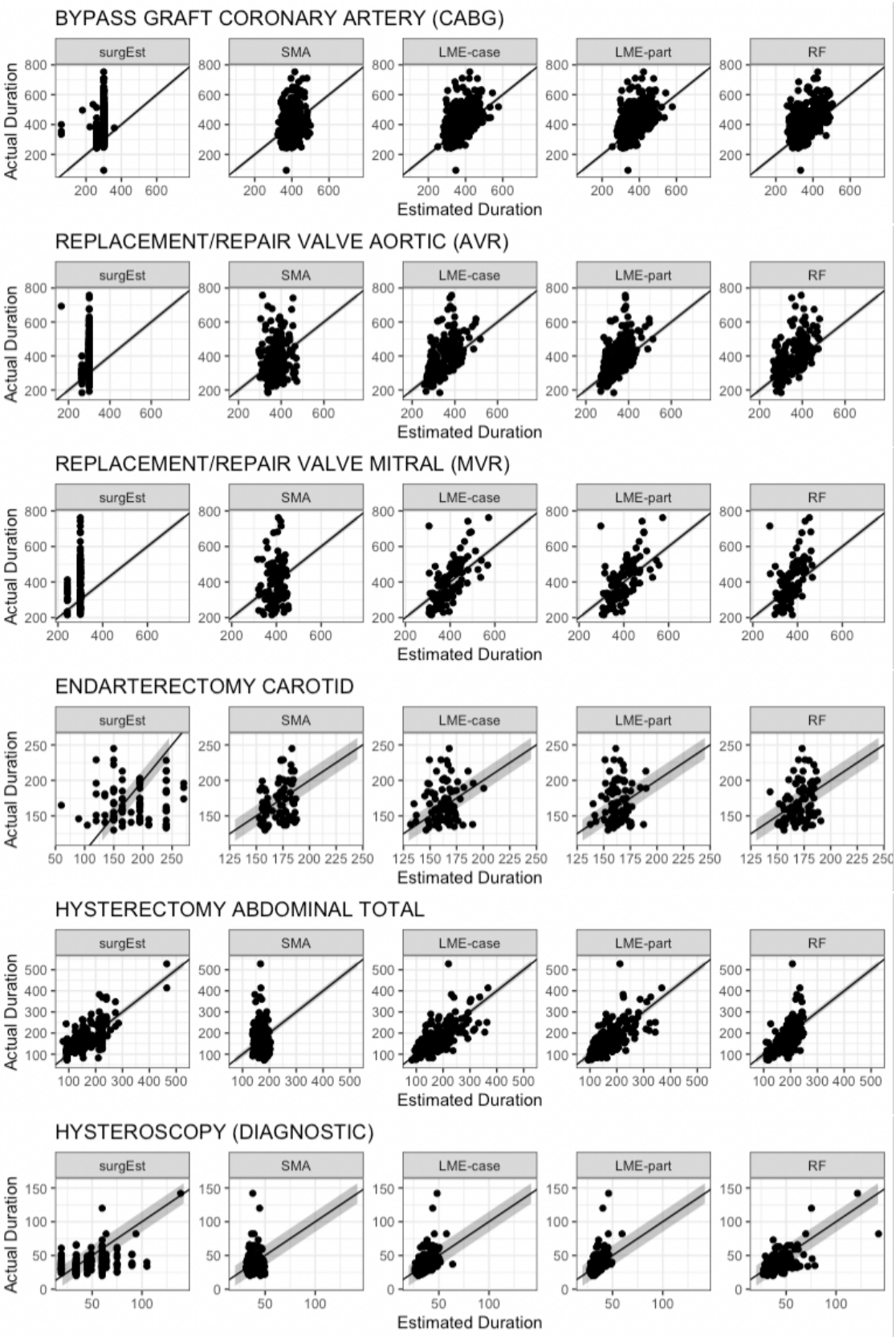


Figure 3-12 Round 2, Model/Procedure: Estimated duration vs actual duration scatter plots

3.3.5. Control (Standardize the Changes, Draw Conclusions)

The next stage would normally be to standardize the changes across all processes, however, due to the limited timeframe and scope in this project this step was not completed. Therefore, additional work would be needed to generate predictive models for all surgical departments and procedures, as well as to create monitoring protocols to ensure the predictions remain accurate and within control measures. It would also be beneficial at this stage to include personnel from all surgical departments to aid in setting control parameters that would be acceptable for each department. This is also the stage where training protocols for those using the new system would be put in place and where details of the new system could be made available to all staff.

Although the changes were not standardized, there are still some plans for standardization and conclusions that can be drawn based on the smaller scale tests. Conclusions include reflections on benefits and challenges faced, as well as future plans.

3.3.5.1. Standardize the Changes

A new method of surgery duration estimation which utilizes past performance data is able to greatly improve accuracy. However, the surgeon estimate is able to further improve the model accuracy, so it should not be eliminated after implementation of a new system. The SMA model or LME-case model could be standardized to work for all surgeries and case services within the OR, provided that at least 19 or more procedures of that type have been completed in the past.

If the SMA model is chosen for implementation, the standardization across procedures for all case services will be quite simple—the same method can be used for them all. Also, there is no need for training a model, only pulling recent surgeries of the same procedure.

If the LME-case model is chosen for implementation, the standardization process may take slightly longer. Multiple models may work better for different case services as they could have

different records and information available. In addition, steps would need to be put in place to ensure that the training data for the model are updated regularly (automatically or based on a specific timeline). Although this process could be more intensive than the SMA model standardization, it is likely to provide more accurate estimates.

To aide in reducing the overtime values associated with the SMA model and to help in controlling for errors in the estimates in both LME-case and SMA models, an adjusted model could be implemented in which surgeons have the ability to override the model estimate if they believe the case will go longer than predicted. Having this ability to override an estimate is a method of control which is currently lacking in the process. In addition, if a model is implemented there is also the option to adjust the model such that the outcomes are shifted. This could be done by adding to or adjusting the estimate in some manner. For example, the SMA model could take 15th highest value of the 19 previous cases—75th percentile—or average plus standard deviation instead of just the average as the next estimate, or the LME-case model could add a constant or standard deviation value to the estimates; this shift would reduce overtime error but would increase the undertime error or vice versa, as desired.

Training in the new method of duration estimates would be needed for anyone using the model. This may include surgeons, surgeon's assistants, and the OR slating clerk. The method of implementing the model also should be done in such a way that it can be easily utilized by any of these individuals and updated as needed. An update of the slating process map would also be needed. Although the duration estimates affect the work of other OR staff (nurses, HCAs, managers), they would not be directly impacted by the change in system and so would not require any additional training, though informing them of the new system could make them better prepared for the change and better able to adapt.

3.3.5.2. Draw Conclusions

The benefits and challenges faced in applying Lean Six Sigma during this project are discussed in detail in Chapter 5. The benefits and challenges specific to assessing and improving the surgery duration estimates is discussed here, along with future plans and areas of interest identified over the course of this project.

The biggest benefit of this project was in recognizing the need for a change in how surgery durations are estimated. Although surgeon estimates are useful in planning slates, studies have shown that they are prone to error and overbooking [86] [87]. Reasons for surgeons to purposely adjust their estimates include the need reduce waiting lists, overestimation of their ability to perform faster than their peers, and pressure to utilize all of their scheduled time in the OR (if time is not used, it may be given to another surgeon). Updating the method used to create these estimates to a data-driven approach removes these biases.

Other benefits of this project—should it be followed through to completion—include the improved accuracy of case and room durations, the ability to standardize how estimates are created across all case services, and improved ability of staff outside of the OR suites to plan for tasks which occur at the start and end of cases. With these benefits in place, significant improvements in other related areas may also be realized due to indirect effects of the improved case estimates. For more details on why it is recommended to complete this project by implementing a new slating method, see Chapter 7.

One challenge lay in determining how the slate and case estimates are currently created. Multiple methods are used between different case services and each surgeon could have a different means of calculating their estimates which could not be tracked at this time. By creating a model, there is now a standardized way in which these surgical durations can be predicted. Not

only does this benefit the process by improving the accuracy of predictions, but it also makes a system that can be more easily monitored, compared, and understood.

Another challenge faced while creating the models was determining which variables to include. Due to predictive models not requiring a rational for inclusion as in explanatory models, all variables could be included. However, there were some variables which could have been useful which were not available. This is an area in which the models may be improved upon in the future. Some potentially beneficial variables which are not currently documented include patient information such as BMI, basic medical history, and a count or severity level of previous surgeries, and staff variables such as team familiarity, surgeon skill and experience, and whether a resident is to be included in the surgery. Another variable that could be useful—similar to surgeon estimate—is a surgeon-based assessment of likelihood for complications to occur.

Once the surgery duration estimates have been improved, there are other areas in which improvements are needed. Turnover times, first case on-time starts, and staff timeliness are all prime areas for improvement. They are areas which directly impact the need for overtime and could greatly benefit from a deep look into the root causes behind inefficiencies. However, after implementing the new duration estimate model, some time may be needed before developing a new baseline in these areas and choosing which to use for the next improvement project as some improvements could already be realized once the slate is more accurately portrayed.

Other areas to consider for improvement in the future are communication and information passing within the OR and updating surgeon preferences. Though these areas are unlikely to provide as much improvement in reducing OR overtime as the prime areas mentioned previously, they do still impact the efficiency and overall functioning of the OR and should be considered once the other projects have completed. Missing or incorrect information could be

affecting turnover times and first case on-time starts, thus may be addressed while implementing improvements in these areas, which is why improving these other areas first would be best.

Updating surgeon preferences is a slow and time intensive task which requires a high level of input from many people (potentially every surgeon) and is unlikely to greatly improve overtime. However, the surgeon preferences should be updated when possible, in order to reduce the need to reset the OR when incorrect. This may be a task that could be assigned to each surgeon to complete over a certain timeframe, though monitoring improvement in this area would require data collection methods to be implemented to track errors due to incorrect surgeon preferences.

Before implementing any of these new projects, thought is needed for how to document errors in these areas. In order to properly improve them, data need to be gathered in such a way that conclusions can be drawn; this means ensuring that every error is tracked and recorded in a standardized way. Employing such data tracking measures and training individuals to use them correctly should occur before starting each new project and may constitute a project of its own.

3.4. Monitor the Outcomes

As with the standardization step of the control stage, this stage was not completed for this project. Once the standardization steps have been completed—including implementing the new estimation process across all departments and creating monitoring and control protocols—protocols must then be followed to ensure that the improvements to estimates are maintained at acceptable levels and within control.

There is already statistical data collected regarding the OR which are compiled and reviewed on a regular basis. The accuracy of duration estimates and rates for over/under-time cases could be included in these reviews in order to monitor the model's performance. This would be the easiest method to monitor and ensure the improvement in duration estimates is maintained.

3.5. Chapter Summary

Within this chapter, the Lean Six Sigma process was used to analyze and improve efficiency within the OR. The project charter was completed (Section 3.1), providing a guideline as to the goals and timeline for the project. The goal for improvement was to reduce the number of late end-of-day cases from 125/month to 30/month. After completing an assessment of the OR workings (Section 3.2) including a breakdown of the different processes within the OR, related departments, and flow of work and materials, the improvement project began (Section 3.3).

Although various areas and projects were considered, improving the accuracy of case duration estimates was determined to provide the best opportunity for improvement. Six procedures were chosen for closer analysis and it was decided that a new data-based method of estimating durations was needed in order to improve on the currently slating method, which used only surgeon estimates. Four models were created—a simple moving average (SMA) model, linear mixed effects models by case duration (LME-case) and by part durations (LME-part), and a random forest (RF) model—and accuracy of the predictions were compared against surgeon estimates. All models performed better than the surgeon estimates, but it was recommended that the SMA model or LME-case model be implemented due to model simplicity and ease of implementation. Due to limitations in the project timeline and scope, the improvement was not standardized across all case services and procedures. At this point, challenges associated with completing the improvement project were assessed and future projects were also discussed.

The final part of the Lean Six Sigma method and this chapter was monitoring the outcomes (Section 3.4). As with standardization of the models across departments and procedures, this step was not completed. However, once implemented, including data about model performance in reviews which are already performed is recommended.

4. RESULTS

The results of the models created for surgery duration prediction were previously assessed (Section 3.3.4.2). In this chapter, the results to be covered are in regard to the Lean Six Sigma process and implementation within the OR environment. By using the Lean Six Sigma method to implement and document this project, many outcomes were produced which will be beneficial for future projects. This includes making it easier to determine what the next steps should be in improving the OR and reducing the amount of work needed when starting new projects as some elements of the OR process analysis are already documented, such as defining CTQs and mapping workflow.

The first outcome of this project was in clarifying the complexity of OR performance and its importance to the hospital, and why Lean Six Sigma and other operational excellence approaches can be useful for implementing lasting improvements. Within the OR environment, some of the outcomes which were achieved include defining the critical to quality (CTQ) requirements and measurables (Figure 3-3), developing a baseline of relevant CTQ measures for this project (Table 3-1), and creating a detailed integrated flowchart portraying OR functions (Figure 3-5). These outputs can be used in future projects to aide in guiding change.

Another outcome of this project which can be used in the future are the recommendations in regard to areas for improvement (Figure 3-6, Table 3-2). With the general assessment of causes, impact, and prioritization levels, it is possible that the analysis could aide in directing efforts to areas that are most likely to generate significant improvements.

Within the process of improving the surgery duration estimates, a key outcome—aside from the models (Table 3-11) and recommendations on which to use and how to standardize (Section 3.3.5.1)—is the identification of other possible root cause problems which could be affecting the

case duration (FMEA, Table 3-7; cause-and-effect diagram, Figure 3-10). Even with the improvement of using a data-driven model to create the duration estimates, it is likely that more work would be needed to meet the goal of reducing late end-of-day surgeries from 125/month to 30/month. Having additional root cause variables already identified will help in establishing the next issue to be tackled in order to meet the goal. An additional benefit of the project was that a short-term solution was applied, which will hopefully have improved surgeon predictions to some degree until a more accurate modeling method can be implemented.

5. DISCUSSION

Based on the results of the project, the use of Lean Six Sigma to assess and improve efficiency in the OR and hospital environment was successful. Although some challenges were faced, the use of the Lean Six Sigma methodology allowed the root problem to be determined and beneficial outcomes to be produced. Along with discussing successes and challenges of this project, this chapter also covers the unique contributions made through this thesis.

5.1. Successes and Challenges

One way in which the Lean Six Sigma methodology was particularly useful was in determining the root problem to be solved. When this project was first proposed, the proposed focus was on first case on-time starts. During the initial stages of defining the initiative (Section 3.1) and characterizing the process (Section 3.2), it was determined that the interest in improving first case on-time starts arose from the belief that it would be a good way to reduce the number of rooms running overtime. However, after analyzing the data it was determined that the first case start times had a limited impact on whether or not the room would end on time. Only 13% of rooms started more than 15 minutes late, while 59% of rooms ended late—and this was without even accounting for cancelled cases. Therefore, the project objective shifted to a broader goal: rather than improving first case on-time starts, the goal was to reduce the number of rooms running overtime at the end of the day, through whatever method was deemed likely to provide the most opportunity for improvement.

Even though the goal was not met by the end of this project, the changes recommended are a good start to ensuring that such a goal can be met in the future. In addition, the analysis of areas in the OR needing improvement provides a solid understanding of the performance gap currently in place, where to work in order to reduce it, and what sort of data need to be collected to do so.

Lean Six Sigma is a cyclical process which allows for continuous improvement, so by continuing to use this method the goal will eventually be met.

As stated previously, the project was not without challenges. The first challenge met was one commonly seen in the OR, which is that hospitals and the OR are a very challenging environment in which to apply Lean Six Sigma due to the complexity of issues and the high level of human factors involved. In fact, the work performed in the OR seems to be almost entirely dependent on human factors: there is no automation, human patients are the "product" being worked on, and variability in surgeries is vast due to variations in both the patients being performed on and the teams which are performing. However, after speaking with individuals from many areas within the OR, it became clear that the complexities which make a Lean Six Sigma project seem difficult also create a great amount of bias in how efficiency is viewed by those within the system. A high-level view of the processes using a data-driven approach is especially useful in order to ensure that the utmost level of improvement can be gained with any negative impacts mitigated for all affected.

Another challenge was working as a newcomer in this environment. Being new to the OR meant that a lot more work was needed to gain a full understanding and appreciation of the system than if staff from within the OR were to complete such a project. A benefit of this though, was that with fresh eyes comes a fresh perspective which is unbiased by years of experience with common problems. Luckily, the team at St. Boniface Hospital was extremely helpful and encouraging, so the learning that was needed was readily provided.

Regardless of the help provided, the experience of the team with Lean Six Sigma and the roles of the team members was another challenge to overcome. Lean Six Sigma projects are normally implemented by a team which includes members with experience both in the area being

improved and with Lean Six Sigma methods (often with training at the Green Belt level), as well as a leader with extensive experience in Lean Six Sigma (most often a certified Lean Six Sigma Black Belt or Master Black Belt). This case was completed as a solo project by a student new to the OR environment and with limited experience implementing Lean Six Sigma projects (training at the Green Belt level but requiring additional experience to become certified). Although the executive team for this project was made up of high-level staff members, they were not directly involved in the assessment or improvement processes undertaken and primarily contributed to the project by giving guidance, answering questions, and providing access to OR staff and data as needed. The results of the project were positive, but it is likely that with a more experienced team greater results could have been achieved, and in less time.

One final challenge which was faced in implementing Lean Six Sigma was with the data. Much of the OR is inter-related, so an error in one area could cause a domino effect which would result in an error in another area. This type of linkage can be difficult to trace and is part of the reason why the tracking of delays, late starts, and long cases were difficult to use; they often had multiple possible root causes for the issue that was listed. Identifying that incorrect case duration estimates were negatively affecting many areas, and that improvements in accuracy could create a more accurately planned approach for staff in the OR, greatly influenced the decision to tackle this problem as the first project. If estimates can be improved and planning of the OR more easily completed, then errors will become more easily tracked and traced to their source. This will be beneficial for future projects.

Regardless of the challenges faced, the Lean Six Sigma project was successful. It is recommended that the Lean Six Sigma cycle be continued with additional projects until the goal is met. The best means of implementing future projects would be to utilize a highly skilled Lean

Six Sigma leader and utilize individuals from within the OR processes, though they will require some training in Lean Six Sigma prior to completing any projects.

5.2. Unique Contributions

The contributions of this thesis to the current body of knowledge can be broken into four categories: (1) acts as a single, comprehensive resource on Lean Six Sigma in healthcare environments, (2) includes extensive detail on the data, tools, and decision-making completed within each step of a Lean Six Sigma improvement project, (3) completed by a unique team compilation compared to common Lean Six Sigma projects, and (4) contributes as a Canada-based study.

This thesis acts as a single resource with a comprehensive background of Lean Six Sigma specific to healthcare, a review of Lean Six Sigma applications in individual OR projects and institutional healthcare environments, and detailed steps for completing a Lean Six Sigma project alongside a sample project completed in one of the most complex hospital environments with data, tools, and decision-making demonstrated. This encompasses all elements needed to aide in educating and influencing doctors, healthcare managers, and hospital operators to endorse a change to using Lean Six Sigma or other continuous improvement methodology for improvement.

In addition, this thesis provides a more robust review of the Lean Six Sigma steps taken over the course of the project than commonly available. While the journal articles reviewed within this work [40]-[73] [84] [85] often stated the type of continuous improvement method (Lean, Six Sigma, Lean Six Sigma, TPS, Kaizen) and steps used (i.e., DMAIC, VSM, RIE, PDSA), they did not provide the same level of detail as to the tools, data, and methods used within those steps. In addition, of the studies reviewed only two [53] [56] assessed the OR in its entirety before

establishing the improvement area, as done in this project; all other studies viewed specific surgeries or tasks. While these narrow interest studies demonstrate how Lean Six Sigma can create a positive change and the approaches used may be generalized for use in other areas, they do not adequately demonstrate how to apply it across an entire system, nor do they show the added benefit gained by using Lean Six Sigma to assess and cyclically continue generating change across a whole institution.

Another area in which this project stands apart from previous research and traditional Lean Six Sigma application is in the makeup of the team. As stated in the previous section, this was a challenge for the project since only the team leader was directly involved in completing the improvement project. In most studies reviewed, multidisciplinary teams were utilized to complete the project, which often included Lean Six Sigma professionals, staff from the area being improved (surgeons, anesthesiologists, nurses, HCAs), engineers, and process owners alongside executive teams for support. The only study with a team remotely similar to this project was completed by Sunder et. al. [71], in which the team did not directly include front-line staff but was rather comprised of a post-graduate student, a professor of Total Quality Management, and a Lean Six Sigma Master Black Belt. Although the inexperience and small size of the team for this project was a challenge, it also demonstrates that—even with minimal active experience and guidance—when trained in its use, Lean Six Sigma is an effective tool for change that can be implemented by anyone willing to learn.

Finally, Canada is not a leader in applying Lean Six Sigma to healthcare. The primary studies assessed within the literature review (Section 2.4.1) took place in the USA (19), Europe (10), India (2), and Brazil (1). The reviews which were assessed (Section 2.4.2) also categorized the locations of studies as primarily being in the USA, Australia, Europe, and Asia, with only 2 of

195 [72] and 2 of 109 [73] empirical articles being Canada-based. Even within the review of Lean and Six Sigma methods in radiology completed by Amaratunga and Dobranowski [88] who were both Canadian-based authors, none of the 23 articles reviewed were based in Canada. Within Manitoba, only three studies were found in which Lean principles were applied in the OR environment; two were related to work done at St. Boniface [84] [85] (as previously noted in Section 2.5) and the other did not fully implement Lean but rather used the Lean tool of VSM to assess patient flow in conjunction with agent-based simulation to improve decision making and efficiency [89]. Therefore, this thesis and the case study project completed contributes to the body of knowledge by being a study completed in a location without extensive publication on or application of the subject.

6. CONCLUSIONS

Although the healthcare sector—and the OR in particular—is a complex environment with copious variation, Lean Six Sigma can be implemented very successfully. In fact, it has already been implemented in many places and shown to generate substantial improvements and savings. The greatest benefit can be seen in places where Lean Six Sigma or other operational excellence mindsets have been implemented across the entire institution, but success has also been found within individual projects such as the one completed here.

One of the reasons that Lean Six Sigma is so successful is that it can be used to see to the root of a problem. In complex environments especially, people from different areas will see different issues, but they may not be able to see how they connect or be able to follow them to their source. The use of data-driven methods to determine root causes can help in finding these connections and differentiating between superficial issues and root causes. In addition, Lean Six Sigma provides a framework which is used not only to identify and improve problem areas, but also to sustain changes and continue improving. This is key to staying efficient in a fast-paced and ever-changing environment, such as in the OR and healthcare as a whole.

For this project, the primary goal identified was to reduce the number of rooms running overtime. All OR processes were assessed for inefficiency and error which would causes rooms to run overtime and six prime areas for improvement were identified. Based on the impact and effort needed to complete projects in these areas, priority was given first to improving case duration estimates, to eventually be followed by projects to improve first case on-time starts, staff timeliness, and turnover times. Communication/information dispersal and updates to surgeon preferences were also identified as needing improvement but were deemed to be of less importance than the other areas.

Within this project, data were used not only to determine what the primary inefficiencies were and which to pursue first, but also as part of the solution. Duration data from past procedures were used to create predictive models which can generate duration estimates for six procedures based on patient and surgical variables. These models do not have the bias that is inherent in surgeon estimates and were found to greatly improve accuracy. More work is needed to implement these models on a wide scale and once the models are implemented, more projects will need to be completed in order to continue reducing overtime rooms until the goal is met. In the end, although there were challenges to completing this project using Lean Six Sigma all were overcome and positive results were realized.

It is hoped that the results of this project will encourage St. Boniface and other hospitals to continue pursuing continuous improvement methodologies such as Lean Six Sigma and expand their use to all areas of the hospital. It has been said that,

"If you're not getting better, you're getting worse."

It is best for everyone that our healthcare continues to get better, and continuous improvement can accomplish this. Improving efficiency in healthcare allows for more people to receive care, reduces wait times, and reduces cost. The possible savings and impact to the quality of life and care for Canadians (or any place that can implement these methods successfully) is great—so what are we waiting for?

7. RECOMMENDATIONS

It is recommended that this project be continued with full implementation of a new predictive slating method. While the changes shown by comparing actual surgical outcomes to the predictive models indicated some of the benefits that could have been achieved (50% reduction in overtime minutes, 65% improvement in sum of error, 27-63% increase in on-time cases), there are some possible impacts from implementing a predictive model and continuing with Lean Six Sigma improvements which—while not quantifiable without wide scale implementation—can be theorized.

Changes in end-of-day late cases could not be assessed, because the benefit of the more accurate estimates were not used when creating the slates for that period and so the schedule was not adjusted to prevent overtime. However, if the models had been in use, then the improvements in estimates would have allowed for staff to be better able to predict not only which rooms would run overtime, but also when each surgery would start and end, thus allowing them to be prepared for completing turnovers, getting the next patient prepped to be ready to go as soon as the room was available, and for the surgical team to be in suite and ready to begin surgery (instead of possibly needing to be called for once the room is prepared).

In addition, morale is likely to improve for a number of reasons. Staff would be able to plan their breaks more easily without worrying that they will be missed because they would be able to predict when they could be spared. The patient and patient's family would be happier because they would have a better idea of duration; family would worry less because there would be fewer cases running overtime, which can be viewed as a bad sign. Fewer cases would need to be cancelled, which would also improve patient satisfaction. Management would be able to plan for rooms which were likely to run overtime, eliminating the stress of having to scramble to find

staff partway through the day when it becomes clear that too many rooms will go overtime to be covered by the change in staff at the end of regularly scheduled surgeries. And finally, staff would be prepared to stay late because they would be informed early and so could adjust plans if needed, ensuring that no one is surprised by unexpected or unwanted overtime.

As well, individuals new to the OR would be better able to understand and succeed in their new environment. Currently, the staff that are best able to successfully plan and manage the OR operations are those who have been there for some time and are very knowledgeable about the inner workings of processes (for example, which rooms are likely to end late due to the surgical team involved or which OR nurses will always have rooms ready on time and their team prepared). This becomes a problem when those individuals retire or leave from their position and new employees enter the workforce as it takes time and experience to develop that level of implicit knowledge. Improving the slate and case duration estimates to match the actual durations is one way in which this learning curve could be mitigated, at least partially.

Another benefit to having accurate duration estimates is in determining how much funding, time, and staff are needed to get through the workload. Proposals to increase time, the number of rooms open, or to hire new staff would benefit from the duration estimates being accurate because they can then be used to show the necessity of such changes and calculate savings that could be realized if such changes were implemented.

If Lean Six Sigma were to be continued, the biggest benefit and impact would come from the cyclical nature of the continuous improvement method. Regardless of how many projects are completed, there is always a next step which can be found and pursued for improvement. In an area such as the OR this is very important because things are constantly changing—new procedures and methods are developed, new technology becomes available, and new people join

the department. Having a methodology in place to ensure that the efficiency of the environment continuously improves along with these changes could greatly benefit the hospital, its staff, and its patients. In addition, if all staff were aware of the basics of Lean Six Sigma and encouraged to contribute, even more improvements could be realized.

In summary, it is highly recommended that this project be completed by implementing a predictive model for slating surgeries and that the use of Lean Six Sigma (or another operational excellence method) be continued. By using such methods, the improvement process is data-driven with quantifiable improvements, standardized solutions, and a plan to ensure that the changes are both sustainable and continuously improved upon. Overall, improving the duration estimates and continuing to improve other areas within the OR will help to reduce costs while also running more smoothly, allowing for more surgeries to be completed while maintaining a positive environment for the staff, patients, and their families.

REFERENCES

[1] E. C. Schneider, D. O. Sarnak , D. Squires, A. Shah and M. M. Doty, "Mirror, Mirror 2017: International Comparison Reflects Flaws and Opportunities for Better U.S. Health Care," The Commonwealth Fund, 2017. [Online]. Available: https://interactives.commonwealthfund.org/2017/july/mirror-mirror/. [Accessed 7 June 2020].

[2] OECD, "Health Spending (indicator)," 2020. [Online]. Available: https://data.oecd.org/healthres/health-spending.htm, doi: 10.1787/8643de7e-en.. [Accessed 3 July 2020].

[3] B. Barua and M. Moir, "Comparing Performance of Universal Health Care Countries, 2019," Fraser Institute, 2019. [Online]. Available: http://www.fraserinstitute.org.

[4] B. Barua and M. Moir, "Waiting Your Turn: Wait Times for Health Care in Canada, 2019 Report," Fraser Institute, 2019. [Online]. Available: http://www.fraserinstitute.org.

[5] B. Barua and D. Jacques, "The Private Cost of Public Queues for Medically Necessary Care, 2019," Fraser Research Bulletin, 2019.

[6] E. Stokes and R. Somerville, "The Economic Costs of Wait Times in Canada," The Centre for Spatial Economics, June 2006. [Online]. Available: https://www.doctorsofbc.ca/sites/default/files/economic_cost_of_wait_times.pdf.

[7] Canadian Institute for Health Information, "Waiting for Health Care in Canada: What We Know and What We Donít Know Waiting for Health Care in Canada: What We

Know and What We Don't Know," 2006. [Online]. Available: https://secure.cihi.ca/free_products/WaitTimesReport_06_e.pdf.

[8] B. Day, "The Consequences of Waiting," in Reducing Wait Times for Health Care: What Canada Can Learn from Theory and International Experience , Fraser Institute, 2013.

[9] A. Picard, "The real challenge to Canada's health system is not wait times," The Globe and Mail, 16 February 2017. [Online]. Available: https://www.theglobeandmail.com/opinion/canada-must-address-the-problem-of-long-waits-for-medical-care/article34056251/.

[10] V. Campbell, Lean Six Sigma "Red X" Green Belt, Winnipeg, Manitoba, 2014.

[11] S. Vig, B. Kadry and A. Macario, "Operating Room Management, Measures of OR Efficiency, and Cost-Effectiveness," in Surgical Patient Care, Switzerland, Springer International Publishing, 2017, pp. 313-325.

[12] J. Divatia and P. Ranganathan, "Can We Improve OR Efficiency?," Journal of Postgraduate Medicine, vol. 61, no. 1, pp. 1-2, 2015.

[13] A. Macario, "Are Your Hospital Operating Rooms "Efficient"?," American Society of Anesthesiologists, vol. 105, no. 2, pp. 237-240, 2006.

[14] R. Kurtz, "7 of the Most Important Metrics for Measuring OR Efficiency," Beckers Hospital Review, 19 January 2012. [Online]. Available: https://www.beckershospitalreview.com/or-efficiencies/7-of-the-most-important-metrics-for-measuring-or-efficiency.html.

[15] T. Foster, "Data for benchmarking your OR's performance," OR Manager, vol. 28, no. 1, pp. 1-5, 2012.

[16] A. J. Fong, M. Smith and A. Langerman, "Efficiency improvement in the operating room," Journal of Surgical Research, vol. 204, pp. 371-383, 2016.

[17] K. Moons, G. Waeyenbergh and L. Pintelon, "Measuring the logistics performance of internal hospital supply chains – a literature study," Omega, pp. 1-13, 2018.

[18] K. J. Saleh, W. M. Novicoff, D. Rion, L. H. MacCracken and R. Siegrist, "Operating Room Throughput: Strategies for Improvement," Journal of Bone and Joint Surgery, vol. 91, pp. 2028-39, 2009.

[19] M. J. Heslin, B. E. Doster, S. L. Daily, M. R. Waldrum, A. M. Boudreaux, A. B. Smith, G. Peters, D. B. Ragan, S. Buchalter, K. I. Bland and L. W. Rue, "Durable Improvements in Efficiency, Safety, and Satisfaction in the Operating Room," American College of Surgeons, vol. 206, no. 5, pp. 1083-89, 2008.

[20] E. Sokolovic, P. Biro, P. Wyss, C. Werthemann, U. Haller, D. Spahn and T. Szucs, "Impact of the reduction of anaesthesia turnover time on operating room efficiency," European Journal of Anaesthesiology, vol. 19, no. 8, pp. 560-63, 2002.

[21] L. Martin and J. Langell, "Improving on-time surgical starts: the impact of implementing pre-OR timeouts and performance pay," Journal of Surgical Research, vol. 219, pp. 222-225, 2017.

[22] C. W. Kimbrough, K. M. McMasters, J. Canary, M. V. Boswell, D. Kim and C. R. Scoggins, "Improved Operating Room Efficiency via Constraint Management:

Experience of a Tertiary-Care Academic Medical Center," Journal of the American College of Surgeons, vol. 221, no. 1, pp. 154-162, 2015.

[23] B. Cardoen, E. Demeulemeester and J. Van der Hoeven, "On the use of planning models in the operating theatre: results of a survey in Flanders," International Journal of Health Planning and Management, vol. 25, pp. 400-414, 2010.

[24] S. Harrison, O. Nugud and H. Benziger, "Operating theatre management: do we lack a mathematician's perspective?," ANZ Journal of Surgery, vol. 83, pp. 5-6, 2013.

[25] J. D. Bronzino, "Engineers as OR team members," AORN Journal, vol. 20, no. 6, pp. 1053-1058, 1974.

[26] M. J. Shaffer and M. D. Shaffer, "The Role of Clinical Engineers in Hospitals: Essential or Expedient?," Hospital Topics, vol. 72, no. 1, pp. 28-35, 2010.

[27] Institute of Medicine Committee on Quality of Health Care in America, Crossing the Quality Chasm: A New Health System for the 21st Century, Washington, D.C.: National Academy Press, 2001.

[28] S. T. Robinson and J. R. Kirsch, "Lean Strategies in the Operating Room," Anesthesiology Clinics, vol. 33, no. 4, pp. 713-730, 2015.

[29] K. Sohrakoff, C. Westlake, E. Key, E. Barth, J. Antognini and V. Johnson, "Optimizing the OR: A Bottom-Up Approach," Hospital Topics, vol. 92, no. 2, pp. 21-27, 2014.

[30] L. Northam, "Data sparks on-time improvements," OR Manager, vol. 26, no. 8, pp. 11-12, 2010.

[31] C. D. Smith, T. Spackman, K. Brommer, M. W. Stewart, M. Vizzini, J. Frye and W. C. Rupp, "Re-Engineering the Operating Room Using Variability Methodology to Improve Health Care Value," Journal of the American College of Surgeons, vol. 216, no. 4, pp. 559-568, 2013.

[32] M. R. Chassin, "Quality of Care: Time to Act," Journal of the American Medical Association, vol. 266, no. 24, pp. 3472-3472, 1991.

[33] M. R. Chassin, "Assessing Strategies for Quality Improvement," Health Affairs, vol. 16, no. 3, p. 151, 1997.

[34] T. Lieberman, "In search of quality health care," Customer Reports, vol. 63, no. 10, p. 35, 1998.

[35] United States Advisory Commission on Consumer Protection and Quality in the Health Care Industry, Quality first : better health care for all Americans : final report to the President of the United States, Washington, D.C.: U.S. G.P.O., 1998.

[36] M. R. Chassin, "Is Health Care Ready for Six Sigma Quality?," The Milbank Quarterly, vol. 76, no. 4, pp. 565-591, 1998.

[37] D. Hill, "Physician Strives to Create Lean, Clean Health Care Machine," The Physician Executive, pp. 62-65, 2001.

[38] C. Caldwell, "A High Quality of Care," Industrial Engineer, pp. 44-48, 2005.

[39] C. Caldwell, J. Brexler and T. Gillem, Lean Six Sigma for Healthcare: A Senior Leader Guide to Improving Cost and Throughput, Milwaukee: ASQ Quality Press, 2005.

[40] K. Burkitt, M. Mor, R. Jain, M. Kruszewski, E. McCray and M. Moreland, "Toyota production system quality improvement initiative improves perioperative antibiotic therapy," American journal of managed care, vol. 15, no. 9, pp. 633-642, 2009.

[41] G. Niemeijer, A. Trip, K. Ahaus, R. Does and K. Wendt, "Quality in trauma care: improving the discharge procedure of patients by means of Lean Six Sigma," Journal of Trauma, vol. 69, no. 3, pp. 614-618, 2010.

[42] M. Iannettoni, W. Lynch, K. Parekh and K. McLaughlin, "Kaizen method for esophagectomy patients: improved quality control, outcomes, and decreased costs," Annals of Thoracic Surgery, vol. 91, no. 4, pp. 1011-1017, 2011.

[43] A. H. Toledo, T. Carroll, E. Arnold, Z. Tulu, T. Caffey, L. E. Kearns and D. A. Gerber, "Reducing liver transplant length of stay: a Lean Six Sigma approach," Progress in Transplantation, vol. 23, no. 4, pp. 350-364, 2013.

[44] G. Niemeijer, E. Flikweert, A. Trip, R. Does, K. Ahaus and A. Boot, "The usefulness of lean six sigma to the development of a clinical pathway for hip fractures," Journal of evaluation in clinical practice, vol. 19, pp. 909-914, 2013.

[45] B. Gayed, S. Black, J. Daggy and I. Munshi, "Redesigning a joint replacement program using Lean Six Sigma in a Veterans Affairs hospital," JAMA Surgery, vol. 148, no. 11, pp. 1050-1056, 2013.

[46] E. Montella, M. V. DiCicco, A. Ferraro, P. Centobelli, E. Raiola, M. Triassi and G. Improta, "The application of Lean Six Sigma methodology to reduce the risk of healthcare–associated infections in surgery departments," Journal of Evaluation in Clinical Practice, vol. 23, p. 530–539, 2017.

[47] Z. Sayeed, A. Anoushiravani, M. El-Othmani, G. Barinaga, Y. Sayeed, P. Cagle and K. J. Saleh, "Implementation of a Hip Fracture Care Pathway Using Lean Six Sigma Methodology in a Level I Trauma Center," Journal of the American Academy of Orthopaedic Surgeons, vol. 26, no. 24, pp. 881-893, 2018.

[48] G. Improta, G. Balato, C. Ricciardi, M. A. Russo, I. Santalucia, M. Triassi and M. Cesarelli, "Lean Six Sigma in healthcare- Fast track surgery for patients undergoing prosthetic hip replacement surgery," The TQM Journal, vol. 31, no. 4, pp. 526-540, 2019.

[49] C. Ricciardi, G. Balato, M. Romano, I. Santalucia, M. Cesarelli and G. Improta, "Fast track surgery for knee replacement surgery: a lean six sigma approach," The TQM Journal, vol. 32, no. 3, pp. 461-474, 2020.

[50] R. Adams, P. Warner, B. Hubbard and T. Goulding, "Decreasing turnaround time between general surgical cases: a six sigma initiative," Journal of nursing administration, vol. 34, no. 3, pp. 140-148, 2004.

[51] R. M. Collar, A. G. Shuman, S. Feiner, A. K. McGonegal, N. Heidel, M. Duck, S. A. McLean, J. E. Billi, D. W. Healy and C. R. Bradford, "Lean Management in Academic Surgery," American College of Surgeons, vol. 214, no. 6, pp. 928-936, 2012.

[52] S. Gleich, M. Nemergut, A. Stans, D. Haile, S. Feigal, A. Heinrich, C. Bosley, J. Ward and S. Tripathi, "Lean Six Sigma handoff process between operating room and pediatric ICU: improvement in patient safety, efficiency and effectiveness," Critical Care, vol. 19, no. 1, p. 523, 2015.

[53] R. Cima, M. Brown, J. Hebl, R. Moore, J. Rogers and A. Kollengode, "Use of lean and six sigma methodology to improve operating room efficiency in a high-volume tertiary-care academic medical center," Journal of the American College of Surgeons, vol. 213, no. 1, pp. 83-92, 2011.

[54] T. L. Stoutzenberger, S. A. Kitner, B. L. Ulrich, I. D'Silva and M. Shah, "Using Lean strategies to improve operating room efficiency," OR Manager, vol. 30, no. 1, pp. 1-4, 2014.

[55] S. Agarwal, J. J. Gallo, A. Parashar, K. K. Agarwal, S. G. Ellis, U. N. Khot, R. Spooner, E. M. Tuzcu and S. R. Kapadia, "Impact of lean six sigma process improvement methodology on cardiac catheterization laboratory efficiency," Cardiovascular Revascularization Medicine, vol. 17, pp. 95-101, 2016.

[56] E. P. Tagge, A. S. Thirumoorthi, J. Lenart, C. Garberoglio and K. W. Mitchell, "Improving operating room efficiency in academic children's hospital using Lean Six Sigma methodology," Journal of Pediatric Surgery, vol. 52, p. 1040–1044, 2017.

[57] C. B. Fairbanks, "Using Six Sigma and Lean Methodologies to Improve OR Throughput," AORN, vol. 86, no. 1, pp. 73-82, 2007.

[58] K. Siddique, S. Elsayed, R. Cheema, S. Mirza and S. Basu, "One-stop cholecystectomy clinic: an application of lean thinking- can it improve the outcomes?," Journal of Perioperative Practice, vol. 22, no. 11, pp. 360-365, 2012.

[59] S. Bhat, E. Gijo and N. Jnanesh, "Application of Lean Six Sigma methodology in the registration process of a hospital," International Journal of Productivity and Performance Management, vol. 63, no. 5, pp. 613-643, 2014.

[60] L. Zhang, K. Runzheimer, E. Bonifer, A. Keulers, E. Piechowiak and A. Mahnken, "Improving Efficiency of Interventional Service by Lean Six Sigma," American College of Radiology, pp. 1200-1203, 2015.

[61] C. Gavriloff, S. Ostrowski-Delahanty and K. Oldfield, "The Impact of Lean Six Sigma Methodology on Patient Scheduling," Nursing Economics, vol. 35, no. 4, pp. 189-193, 2017.

[62] W. J. Glover, E. M. Van Aken, K. Creehan and J. Skevington, "Case Study on Using Lean Principles to Improve Turnaround Time and First Case Starts in an Operating Room," Society for Health Systems Conference and Expo, 2009.

[63] P. Ramos, E. Bonfá, P. Goulart, M. Medeiros, N. Cruz, P. Puceh-Leão and B. Feiner, "First-case tardiness reduction in a tertiary academic medical center operating room: A lean six sigma perspective," Perioperative Care and Operating Room Management, vol. 5, pp. 7-12, 2016.

[64] R. Deldar, T. Soleimani, C. Harmon, L. H. Stevens, R. Sood, S. S. Tholpady and M. W. Chu, "Improving first case start times using Lean in an academic medical center," American Journal of Surgery, vol. 213, pp. 991-995, 2017.

[65] J. Franklin and T. Franklin, "Improving Preoperative Throughput," American Society of PeriAnesthesia Nurses, vol. 32, no. 1, pp. 38-44, 2017.

[66] H. Frankel, W. Crede, J. Topal, S. Roumanis, M. Devlin and A. Foley, "Use of corporate Six Sigma performance- improvement strategies to reduce incidence of catheter- related bloodstream infections in a surgical ICU," Journal of the American College of Surgeons, vol. 201, no. 3, pp. 349-358, 2005.

[67] R. Muder, C. Cunningham, E. McCray, C. Quier, P. Perreiah and R. Jain, "Implementation of an industrial systems-engineering approach to reduce the incidence of methicillin-resistant Staphylococcus aureus infection," Infection control and hospital epidemiology, vol. 29, no. 8, pp. 702-708, 2008.

[68] G. Improta, M. Cesarelli, P. Montuori, L. C. Santillo and M. Triassi, "Reducing the risk of healthcare-associated infections through Lean Six Sigma: The case of the medicine areas at the Federico II University Hospital in Naples (Italy)," Journal of Evaluation in Clinical Practice, vol. 24, p. 338–346, 2018.

[69] P. Schwarz, K. Pannes, M. Nathan, H. Reimer, A. Kleespies and N. Kuhn, "Lean processes for optimizing OR capacity utilization: prospective analysis before and after implementation of value stream mapping (VSM)," Langenbeck's Archives of Surgery, vol. 396, no. 7, pp. 1047-1053, 2011.

[70] J.-M. Lindholm, I. Laine, H. Hippala, P. Ylinen and R. Tuuminen, "Improving eye care services with a lean approach," Acta Ophthalmologica, pp. 1-5, 2018.

[71] V. Sunder M, S. Mahalingam and S. N. Krishna M, "Improving patients' satisfaction in a mobile hospital using Lean Six Sigma – a design-thinking intervention," Production Planning & Control, vol. 31, no. 6, pp. 512-526, 2020.

[72] S. Deblois and L. Lepanto, "Lean and Six Sigma in acute care: a systematic review of reviews," International Journal of Health Care Quality Assurance, vol. 29, no. 2, pp. 192-208, 2016.

[73] A. D'Andreamatteoa, L. Iannia, F. Legab and M. Sargiacomo, "Lean in healthcare: A comprehensive review," Health Policy, vol. 119, pp. 1197-1209, 2015.

[74] F. Worku, "Using Lean Principles to Develop the Virginia Mason Production System (VMPS)," Colleaga, 2020. [Online]. Available: https://www.colleaga.org/case-study/using-lean-principles-develop-virginia-mason-production-system-vmps. [Accessed June 2020].

[75] S. J. Swensen, J. A. Dilling, D. S. Milliner, R. S. Zimmerman, W. J. Maples, M. E. Lindsay and G. B. Bartley, "Quality: The Mayo Clinic Approach," American Journal of Medical Quality , vol. 24, pp. 428-440, 2009.

[76] N. I. Comfere, J. C. Matulis and J. C. O'Horo, "Quality improvement and healthcare: The Mayo Clinic quality Academy experience," Journal of Clinical Tuberculosis Other Mycobacterial Diseases, vol. 100170, pp. 1-5, 2020.

[77] U.S. News & World Report, "Hospital Rankings & Ratings," 2020. [Online]. Available: https://health.usnews.com/best-hospitals. [Accessed June 2020].

[78] Cleveland Clinic, "Cleveland Clinic Improvement Model (CCIM)," 2018. [Online]. Available: https://my.clevelandclinic.org/-/scassets/files/org/clinical-transformation/cc-improvement-model.ashx?la=en. [Accessed June 2020].

[79] Cleveland Clinic, "Continuous Improvement: Overview," 2020. [Online]. Available: https://my.clevelandclinic.org/departments/patient-experience/depts/continuous-improvement. [Accessed June 2020].

[80] K. Barnas, "ThedaCare's Business Performance System: Sustaining Continuous Daily Improvement Through Hospital Management in a Lean Environment," The Joint Commission Journal on Quality and Patient Safety, vol. 37, no. 9, pp. 387-399, 2011.

[81] M. Mannon, "Lean Healthcare and Quality Management: The Experience of ThedaCare," Quality Management Journal, vol. 21, no. 1, pp. 7-10, 2014.

[82] D. O. Weber, "Toyota-style Management Drives Virginia Mason," The Physician Executive, pp. 12-17, 2006.

[83] D. B. Henrique and M. G. Filho, "A systematic literature review of empirical research in Lean and Six Sigma in healthcare," Total Quality Management & Business Excellence, vol. 31, no. 4, pp. 429-449, 2020.

[84] Z. Piggott, E. Weldon, T. Strome and A. Chochinov, "Application of Lean principles to improve early cardiac care in the emergency department," CJEM, vol. 13, no. 5, pp. 325-332, 2011.

[85] C. Swanson, "Reorganizing a Resuscitation Room Using Six Sigma (6S) Principles," J Emerg Nurs, vol. 40, no. 4, pp. 371-376, 2014.

[86] A. Fugener, S. Schiffels and R. Kolisch, "Overutilization and underutilization of operating rooms - insights from behavioral health care operations management," Health Care Manag Sci, vol. 20, p. 115–128, 2017.

[87] J. J. Pandit and A. Carey, "Estimating the duration of common elective operations: implications for operating list management," Anesthesia, vol. 61, pp. 768-776, 2006.

[88] T. Amaratunga and J. Dobranowski, "Systematic Review of the Application of Lean and Six Sigma Quality Improvement Methodologies in Radiology," Journal of the American College of Radiology, vol. 13, no. 9, pp. 1088-1095, 2016.

[89] Y. Xie and Q. Peng, "Integration of value stream mapping and agent-based modeling for OR improvement," Business Process Management Journal, vol. 18, no. 4, pp. 585-599, 2012.

APPENDICES

LIST OF APPENDIX FIGURES

LIST OF APPENDIX TABLES

APPENDIX A **LEAN SIX SIGMA BACKGROUND**

Quality in product manufacturing has a simple beginning; with guilds and craftsmen honing their trades to provide products that met their clients' standards. However, with the industrial revolution in the 1800s came a change in focus to relying on the skill of the workers and product inspection for quality, as well as a push for improvement in efficiency. Quality control and management became a major focus during World War II as it was key for supplying the war effort and ensuring safety. The quality revolution began in Japan as they needed to rebuild after the war, after which the concepts were adapted and widely used in the Americas in response to Japan's success. The individuals who were key to the development of quality control and the principles behind their philosophies are discussed below, along with how they turned into the current methodologies of Lean and Six Sigma.

A.1 Background

One of the foundations of quality control began with development of key statistical analysis principles in the early 1800s by Charles Gauss, namely the concept of normal distributions which is also known by the shape it creates as a bell-shaped curve.

With the transition to factories and the need for inspection, came a new approach to management which was developed by Frederick W. Taylor in the late 1800s [1]. He believed that how hard the employees worked was less important than optimizing the process used to complete the work and that all workers were motivated by money, so those who were more productive should be rewarded with increased pay. Based on his studies on efficiency in the workplace, Taylor developed the philosophy of Scientific Management, which has four principles: (1) use the scientific method to study work and determine the most efficient method for completing tasks, (2) match workers to jobs based on their capabilities and motivation, and

train them in efficient practices, (3) monitor performance to ensure that the efficient methods are used, and (4) allocate work such that managers plan and train, and workers perform their tasks efficiently.

This focus on efficiency was continued by Henry Ford in the manufacturing of the Model T cars in 1913 [2]. He managed to create a process which was efficient by eliminating waste and producing flow using an assembly line (which had been implemented for the first time by E.B. Whitney in 1798, in his uniformity system). This was the first system to accomplish mass production of a standardized product. Some of Ford's innovations which are used in the current Lean Six Sigma methodology include standardization, waste minimization, just-in-time manufacturing, and customer service. However, the system implemented by Ford was limited in that it was unable to allow for variation and had demeaning job structures which were not sustainable long-term.

Following this in the 1920s, Walter A. Shewhart developed statistical methods to improve quality control, including control charts to identify different classes of variation, and was the first to show that a process is in need of correction when it deviates three-sigma from the mean—a key component of the current Six Sigma philosophy [3]. He also created the PDCA improvement cycle, which stands for Plan, Do, Check, Act (or Adjust). Shewhart acted as mentor for the following two individuals who contributed to the development of Six Sigma into the 1950s, W. Edwards Deming and Joseph M. Juran.

Deming was a disciple of Shewhart's teachings on quality control and known for spreading the teachings of his mentor and promoting the concept of continuous improvement and the PDCA cycle for assessing process problems [4]. From there, Deming also built on the concepts of his mentor to create his own philosophy called the "System of Profound Knowledge", which

takes the approach of viewing organizations as interdependent systems [4] [5]. The four key parts to this philosophy are: (1) appreciation for a system, (2) knowledge of variation, (3) theory of knowledge (PDCA), and (4) psychology of change. He is also known for his "14 Points for the Transformation of Management", "The Seven Deadly Diseases of Management", and the "Deming Chain Reaction". After WWII, Deming went to Japan where he lectured and trained thousands of managers and engineers for many years and is widely credited for helping Japan to transform to a powerhouse of manufacturing after the devastation of the war, which is also when America began to take note of Deming's teachings.

Juran's contributions came in promoting the impact of top and middle management on quality control and improvement, and the human relations as they contribute to quality [6]. He created the "Juran Trilogy" of three managerial processes: quality planning, quality control, and quality improvement. He also brought to light the appropriateness of the Pareto principle, also known as the 80:20 principle, being applied to quality issues; for example, that 80% of a problem is caused by 20% of the causes or that 20% of workers make up 80% of all results. Juran led seminars in Japan for years before founding the Juran Institute to continue spreading the concepts of quality control and improvement.

Another key contributor to the quality control movement was Armand V. Feigenbaum, who was first to coin the term Total Quality Control, now more widely known as Total Quality Management [7]. His contributions focused on the relationship of financial performance to quality and the extension of quality control practices from product manufacturing into all areas of business. He introduced the concept of the "hidden plant"; the idea that up to 40% of a manufacturing plant's capacity was used to fix errors. Feigenbaum wrote a book, originally published in 1951, which was re-released on its third edition under the title "Total Quality

Control" around the same time that Deming's work in Japan was gaining traction in America in the 1980s. These coinciding events caused new interest to be given to his work and principles.

After these influencers on quality, came Kaoru Ishikawa's contributions in the 1980s [7]. Ishikawa aided strongly in developing quality initiatives in Japan—using the base provided by the teachings of Juran and Deming—and is known for creating the fishbone diagram (also known as the cause-and-effect or Ishikawa diagram) and quality circles which are groups within companies which meet regularly to tackle quality issues. Ishikawa called the Japanese method of quality management Companywide Quality Control in order to differentiate it from the American approaches to quality control, such as Total Quality Control.

The philosophies of Genichi Taguchi of Japan became well-known in American businesses in the 1980s as well [8] [9]. Taguchi created the Taguchi Loss Function and Taguchi Robust Design, both which deal with the issue of variability and customer satisfaction. Traditionally, upper and lower tolerances were used to ensure that products stayed within a specified range of values, with an ideal value being between the two, and any product within the tolerances would be considered acceptable. However, the Taguchi Loss Function improves on the traditional tolerance-based approach and essentially states that any deviation from the ideal will result in some level of dissatisfaction; the larger the deviation, the greater the dissatisfaction until the product surpasses the tolerance levels, at which point the product is no longer viable (reaching a plateaued level of dissatisfaction). Taguchi Robust Design focuses on ensuring that a product can be produced with minimal variation by adjusting control factors such that the system is less sensitive to variability due to uncontrollable factors, or "noise".

Philip B. Crosby is well-known for creating the concept of "zero-defects" in the 1960s and wrote multiple books on quality in the 1980s and 1990s which were easily consumable and thus

made quality control more accessible to the average population [10]. His “zero-defects” concept focused on prevention, rather than inspection or correction, of errors. He believed that conformance to requirements was the measure of quality rather than the ambiguous labels of “good/bad” or “high/low”. All of his philosophies centred around his Four Absolutes of Quality Management, which are: (1) quality means conformance to requirements, not goodness; (2) quality is achieved by prevention, not appraisal; (3) quality has a performance standard of zero defects, not acceptable quality levels; and (4) quality is measured by the price of nonconformance, not indexes. Crosby’s philosophies also included “14 Steps to Quality Improvement” and “Five Characteristics of an Eternally Successful Organization”.

A.2 Lean Manufacturing

Lean manufacturing was created by Taiichi Ohno and Shigeo Shingo at the Toyota Motor Company between 1949-1975, who called their process improvement system the Toyota Production System (TPS). The coining of the term “Lean” came from John Krafcik in 1987, who was a researcher at MIT at the time. Elements of the teachings and philosophies from Ford, Juran, Deming, and Ishikawa were all used to create Lean manufacturing. Fujio Cho, Toyota’s former president, created “the Toyota Way” in 2001, which detailed the organizational culture in order to compensate for inconsistencies in understanding of daily management principles among Toyota managers [11]. Simplified versions of the TPS and Toyota Way models are shown in Figure A-1, showing the pillars and foundation which hold up each. Although TPS continues to evolve as the needs of the company change, these core pillars and foundations remain constant.

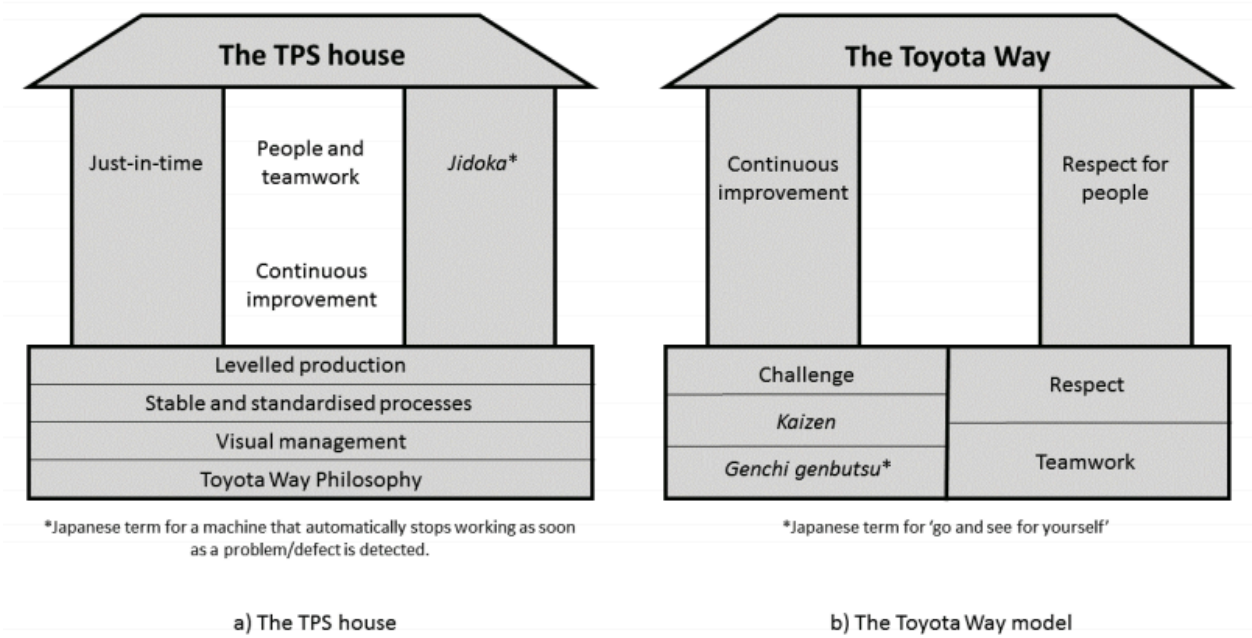


Figure A-1 The TPS house and the Toyota Way model [12]

The element of Kaizen in Lean was popularized by Masaaki Imai in 1985 [13]. Kaizen, in modern usage, is used for quick, focused improvement projects, usually taking place over the course of a week. Kaizens follow 10 principles and produce significant, fast results which can be used to promote continued growth and improvement by increasing the enthusiasm and support of the workers.

Although Lean was initially introduced as synonymous with TPS in 1991, with the book "The Machine That Changed the World" by James P. Womack, Daniel T. Jones, and Daniel Roos [14], the concepts behind Lean have varied since. Lean thinking, as defined by Womack and Jones in their 1996 book "Lean Thinking: Banish Waste and Create Wealth in Your Corporation" [15], consisted of five principles: (1) specify value, (2) identify/map the value stream, (3) create flow, (4) use a pull system, and (5) pursue perfection. The key concepts behind Lean work to improve flow and reduce the 8 types of waste in processes: defects, overproduction, waiting, unused talent, transportation, inventory, motion, and excess processing.

This varies from TPS in that Lean is not reliant on certain elements which are integral to TPS, such as Jidoka, Kaizen, and human/employee relations [16].

A.3 Six Sigma

Six Sigma began in Motorola in 1987 with the coining of the term "Six Sigma" by Motorola engineer William Smith, with the support of Robert Galvin, who was Motorola's CEO at the time, and Mikel Harry, who was a key architect behind the Six Sigma methods. The name comes from the standard deviation needed to reduce variation in the process to 6 standard deviations from the ideal (99.9996% accuracy or 3.4 defects per million). Many of the Six Sigma principles and tools are taken from those whose quality control philosophies came before (Gauss, Ford, Shewhart, Juran, Deming, Ishikawa, Taguchi, Crosby, TPS/Kaizen). The key concepts which drive Six Sigma are defect elimination, reduction in variability, and root cause analysis [17] [18]. A variation on Shewhart's PDCA model is the foundation for Six Sigma, called DMAIC (Define, Measure, Analyze, Improve, Control). After Motorola achieved success using their Six Sigma method, other businesses in America began to use the improvement methodology themselves, including Texas Instruments, ABB, and Kodak. General Electric and Allied Signal brought Six Sigma to the forefront in the mid-1990s by crediting Six Sigma with their increase in market capitalization.

A.4 Lean Six Sigma

Today, most companies concerned with quality control and improvement have at least one of the methods discussed above currently in place or a variation of them. Lean and Six Sigma methodologies, when used together, have a complementary effect and were first incorporated into a single philosophy in the early 2000s, with the first instance being credited to "Leaning into Six Sigma" by Barbara Wheat, Chuck Mills, and Mike Carnell [19]. Alone, Six Sigma will miss

improving process flow, while the Lean principles are lacking the statistical tools needed to fully assess the process capabilities; together, these two methodologies provide a much more comprehensive improvement philosophy. Thus, by first implementing the Lean concepts to reduce waste and improve cycle time, then applying Six Sigma to reduce process variation and eliminate the root causes of defects, a system can be made to be the most robust and efficient as possible. Including elements such as Kaizen (quick improvements) can further improve the effectiveness of the methodology [20]. Another change to Lean Six Sigma was the shift from use primarily in manufacturing to widespread use in other industries as well, such as healthcare, supply chain, administration, and customer service.

APPENDIX B **PROJECT CHARTER**

Team: St. Boniface Hospital
Date: 05/16/2018
Project name: OR On-Time End-of-Day

Project Background:

Operational Excellence is a mindset of continuous improvement by set methods. Two types of methodologies used are Lean manufacturing and Six Sigma. These methods were originally developed in manufacturing settings; Lean manufacturing in Toyota Motor Company and Six Sigma in Motorola. The focus of Lean is to produce the most value for the customer with fewer resources while Six Sigma focuses on the elimination of defects and reduction in cycle time. Both methods strive to eliminate waste in order to reach these outcomes.

Although these methods originated in manufacturing, they have expanded into all types of service industries with great success. Even highly variable processes, such as those in a hospital operating room (OR), still have elements which can be standardized and improved. For this project the St. Boniface OR efficiency is to be assessed, with an end goal being to consistently have on-time end-of-day in the OR.

Business Case:

From April 2017 to April 2018, 35% of elective surgeries began more than 15 minutes late and 38% ended more than 15 minutes late. In addition, an average of 110 end-of-day cases per month finished over 15 minutes late, accounting for 51.5% of all last cases. This resulted in 2152 hours of overtime.

Aside from the monetary cost of overtime, additional negative effects of late cases and end-of-day in the OR include poor patient flow, cancelled end-of-day cases, resource utilization issues, and poor patient experience. These are problems not only in terms of cost for the hospital, but also for the health and safety of patients.

Improving OR end-of-day times will reduce the number of cancelled cases and overtime required, thus improving patient care and reducing costs.

Project Objective:

The primary objective of this project is to develop and implement a strategy to improve the consistency with which the rooms for elective slated surgeries finish on time. The goal is to provide recommendation to reduce the number of late end-of-day surgeries (finishing past the scheduled end time) from 125/month to an average of 30/month.

Scope/Boundaries:

The scope of this project includes all processes required to run the OR, including, but not limited to, L2PO, OR room preparation, room turnover, start and end of day tasks, and slating.

In scope:

- Elective surgeries
- Emergency surgeries as they impact elective surgeries and staffing
- Slating times and procedures
- Start time for staff
- Staff (surgeons, anesthetists, CRNs, nurses, HCA, etc.)
- All processes (L2PO, turnovers, OR communication, OR prep, case cart prep, labs, etc.)
    - Implementation
    - Standardization
    - Task allocation

Out of scope

- Surgery procedure
- Total work time for staff
- Required components for processes (i.e., steps involved in turning over a room)
- Emergency surgery cases
- Cases which are no longer done at St. Boniface Hospital (cystoscopy flexible, primarily)

Schedule and Milestones:

Start	End	Duration	Milestone
Apr 23, 2018	Apr 27, 2018	5 days	Gemba – OR walkthrough
May 11, 2018	May 18, 2018	1 week	Complete project charter
Apr 27, 2018	May 25, 2018	4 weeks	Understanding the Process (mapping, etc.)
May 25, 2018	Jun 1, 2018	1 week	1. Establish the focus
Jun 1, 2018	Sep 1, 2018	3 months	2. Examine the current situation
Sep 1, 2018	Dec 1, 2018	3 months	3. Analyze the causes
Dec 1, 2018	Sep 1, 2019	9 months	4. Act on the causes
Jul 16, 2019	Aug 20, 2019	5 weeks	Write mid-project report
Sep 1, 2019	Dec 31, 2019	4 months	5. Study the results
Jan 1, 2020	Aug 1, 2020	7 months	6. Standardize the changes
Aug 1, 2020	Aug 31, 2020	1 month	7. Draw conclusions
Aug 10, 2020	Aug 31, 2020	3 weeks	Write final project report

Note: milestones numbered 1-7 are the steps involved in a Lean Six Sigma improvement project

Team/Resources:

Team:

Name, title	Role	Responsibilities
Lance Barber Director of Surgery	Project champion/sponsor	Guidance, executive team
Essi Shams Quality Improvement and Patient Safety	Quality and Productivity	Guidance
Kurt Shaw Director of Transformation	Op Ex expert	Guidance
Sarah Slagerman* UofM MSc. in Engineering student	Team leader	Organize team/project, data analysis
Dawn Affleck Program Team Manager, OR	Process Owner	Process knowledge, executive team
Lorena Thiesson Program Team Manager, PAC/B2/L2/NFA	Process Owner	Process knowledge, executive team
Tamara Miller Site Leader, Department of Anesthesia, Perioperative and Pain Medicine	Process Owner	Process knowledge, executive team
Mohamed Yusuf/Scott Vandale Projects and Systems Coordinator	Data consultant	Data acquisition, executive team
Tracy Ptak Business Improvement Finance Officer	Data consultant	Data acquisition
Kathy Ott Improvement Coach, Transformation	Quality and Productivity	Guidance, resources

*Note: Unlike a project conducted in a well-developed Lean Six Sigma environment, this project was completed as a trial and therefore did not include the participation of a Lean Six Sigma Green/Black Belt team member. Sarah Slagerman, as the team leader and completing this project as a component of her Master of Science in Engineering, acted as the primary source of Lean Six Sigma knowledge and driving force of the project.

Other resources:

- CMO, Chief Medical Officer
- CNO, Chief Nursing Officer

Budget and expenses:

- There is currently no allocation of funds budgeted for this project.

Risks:

There are a number of risks identified in moving forward with this project:

1. Vacations
2. Running over schedule
3. Mohamed's absence, beginning Jun. 1; Scott's replacement, beginning Dec. 2018
4. Reaction of staff to changes and methods

Risk Mitigation/Action Plans:

1. Vacations
   Sufficient time has been allocated to each section of the project that a week of vacation would not be an issue in moving forward with any step of the project. In addition, if any individuals on the project team are taking vacation time, they will leave an alternative contact that may speak/attend meetings/take on their responsibilities in their absence.

2. Falling behind schedule
   If the project is deemed to be falling behind schedule, a number of possible actions may be taken to bring the project back on track. This includes implementing weekly team meetings, increasing communication with the Op Ex Expert and Process Owners, and seeking additional team members from the OR and L2PO teams for aid in tasks.

3. Mohamed's leave of absence, beginning June 1; Scott's replacement, beginning Dec. 2018
   As Mohamed's primary role in this project is as the data consultant, in preparation for his leave of absence the majority of data required for the project have already been provided to the team leader. While he is away, if additional access to data is required Tracy Ptak will be the primary contact. Additional time will be given when requesting data during this time to allow for any delays in data acquisition due to the alternate contact.
   *Note: once Scott was hired, he was able to take over Mohamed's role in the project.

4. Reaction of staff to changes and methods
   To ensure staff will be receptive to the project and resulting changes implemented, there are a number of action plans in place to ensure acceptance and successful implementation. Firstly, a report detailing the project findings and data will be made available. Secondly, a change management process will be incorporated into the implementation plan so that all individuals affected by the changes will make successful transitions. This will involve including affected individuals in the problem-solving process, creating a compelling argument supporting the changes, and providing support and coaching as necessary.

Document Revision Control:

1. Changes to the terms of reference document that materially affect the intent shall be approved by the Committee and be issued with a revision number.
2. An administrative change, such as grammar change or an update that does not materially change this document may be revised without a revision number. The project leader/owner will have the discretion whether the minor revision requires to be issued.

Revision Date	Revision Number	Description
May 16, 2018	1	First draft
July 10, 2018	2	Added cystoscopy flexible cases to out-of-scope
July 25, 2019	3	Added Scott Vandale to the project
Mar 15, 2020	4	Adjusted schedule due to COVID-19

Project Sign-off:

_____

Lance Barber, Director of Surgery

_____

Sarah Slagerman, Team Leader

APPENDIX C **DATA SORTING**

Table C-1 St. Boniface Hospital surgical data headings from April 2017 to April 2018

Variable	Data type	Calculation
CaseNum	Case info	
Case Date	Case info	
Room Description	Case info	
DOW	Case info	
Case Service	Case info	
Primary Procedure Description	Case info	
Scheduled Main Surgeon	Case info	
Scheduled Anesthesiologist	Case info	
Patient Status	Patient info	
Anesthesia Type	Case info	
ASA Score	Patient info	
Age – Surgery Date	Patient info	
Sex	Patient info	
Scheduled Start Value	Time value (estimate)	
Scheduled Stop Value	Time value (estimate)	
Patient In Value	Time value (actual)	
Anesthesia Value	Time value (actual)	
Proc Start Value	Time value (actual)	
Proc Stop Value	Time value (actual)	
Patient Out Value	Time value (actual)	
Sched Duration	Duration value (estimate)	Sched. stop – Sched. start value
Prep dur	Duration value (actual)	Anesthesia – Patient in value
Anes dur	Duration value (actual)	Proc start – Anesthesia value
Proc Duration	Duration value (actual)	Proc stop – Proc start value
Wrapup dur	Duration value (actual)	Patient out – Proc stop value
Case Duration	Duration value (actual)	Patient out – Patient in value
Planned order	Daily schedule info	Sorted based on Scheduled Start Value, Date, Room
Actual order	Daily schedule info	Sorted based on Patient In Value, Date, Room
Prev Sched Stop	Time value (estimate)	Found based on previous case Scheduled Stop Value, Date, and Room
Prev Patient Out Time	Time value (actual)	Found based on previous case Patient Out Value, Date, and Room
Next Patient In Time	Time value (actual)	Found based on next case Patient In Value, Date, and Room
Turnover dur	Duration value (actual)	Patient Out Value – Next Patient In Time
First Scheduled Start	Daily schedule info	Found based on Planned Order, Date, Room
Last Scheduled Start	Daily schedule info	Found based on Planned Order, Date, Room
isFirst	Daily schedule info	True/False based on First Sched Start, Sched Start Value
isLast	Daily schedule info	True/False based on Last Sched Start, Sched Stop Value
Num of Procedures	Case info	Calculated based on Secondary Procedures with same CaseNum

In order to use the data, they first had to be filtered and sorted. The steps taken to do so are shown in Table C-2. After determining which surgeries were to be included for more in-depth assessment (for which the process for choosing is detailed in APPENDIX E) the data were filtered for those six surgeries as well.

Table C-2 Data filtering steps

Action	Remove	Remaining
Start		9805
Filter for elective cases	2279	7526
Filter out duplicates (multiple secondary procedures completed at same time)	497	7029
Filter out surgeries "cystoscopy stent removal", "cystoscopy flexible" & room "SBH ER"	791	6238
Eliminate additional duplicate cases of "328417", "386867"	3	6235
Filter for the six chosen surgeries:	5098	1137
Bypass graft coronary artery (CABG)		*477*
Replacement/repair valve aortic (AVR)		*164*
Replacement/repair valve mitral (MVR)		*82*
Endarterectomy carotid (END)		*73*
Hysterectomy abdominal total (HAT)		*159*
Hysteroscopy, diagnostic (HYSD)		*182*
Eliminate cases with duration errors (details in Appendix G.2):	17	1120
Bypass graft coronary artery (CABG)	*8*	*469*
Replacement/repair valve aortic (AVR)	*5*	*159*
Replacement/repair valve mitral (MVR)	*1*	*81*
Endarterectomy carotid (END)	*2*	*71*
Hysterectomy abdominal total (HAT)	*0*	*159*
Hysteroscopy, diagnostic (HYSD)	1	*181*
Eliminate cases that are outliers (details in Appendix G.2):	15+5+27	1073
Bypass graft coronary artery (CABG)	*20*	*462*
Replacement/repair valve aortic (AVR)	*7*	*156*
Replacement/repair valve mitral (MVR)	*3*	*81*
Endarterectomy carotid (END)	*1*	*71*
Hysterectomy abdominal total (HAT)	*4*	*157*
Hysteroscopy, diagnostic (HYSD)	12	*178*

APPENDIX D **ORIGINAL NOTES AND MAPPING**

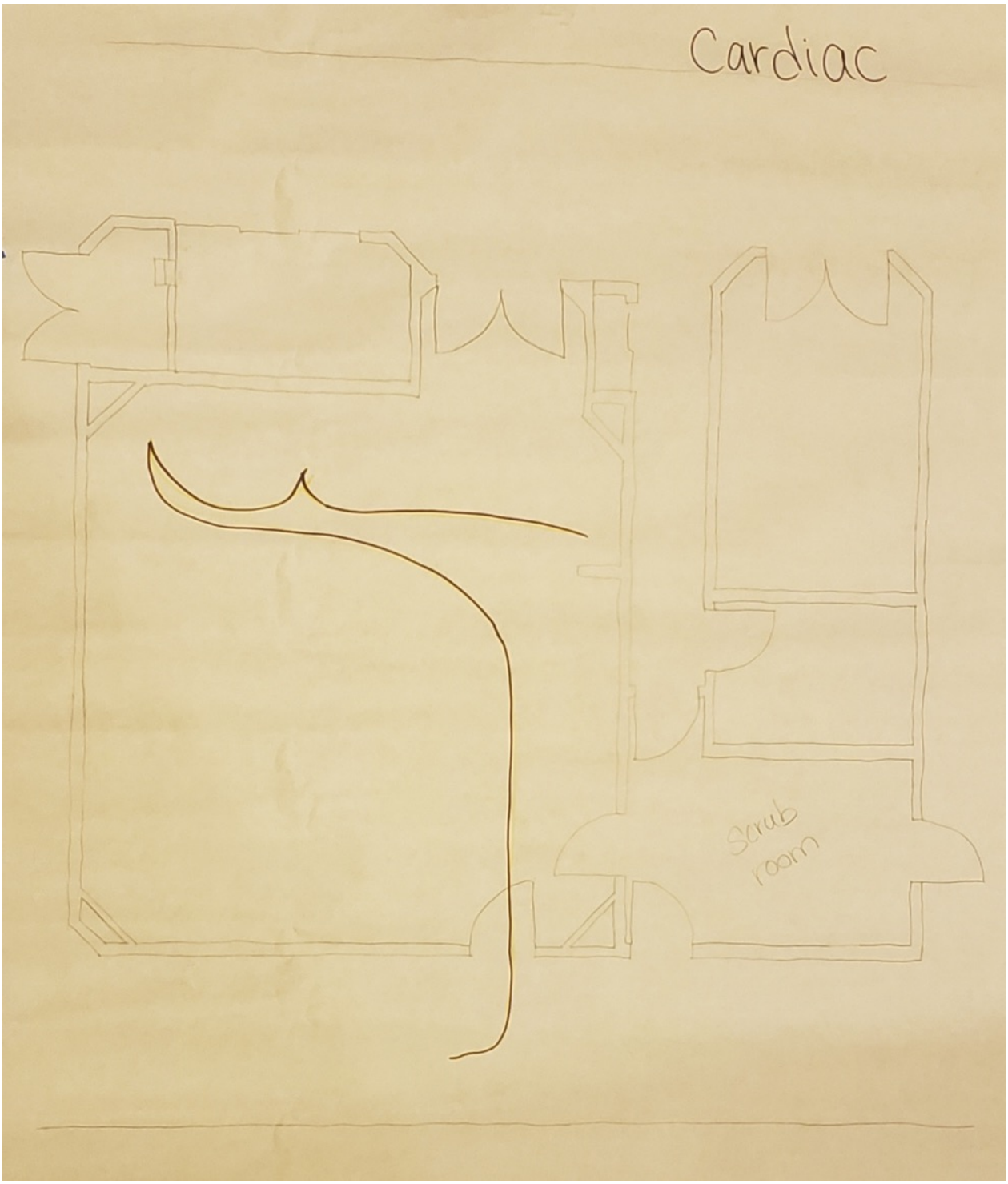


Figure D-1 Cardiac room spaghetti map

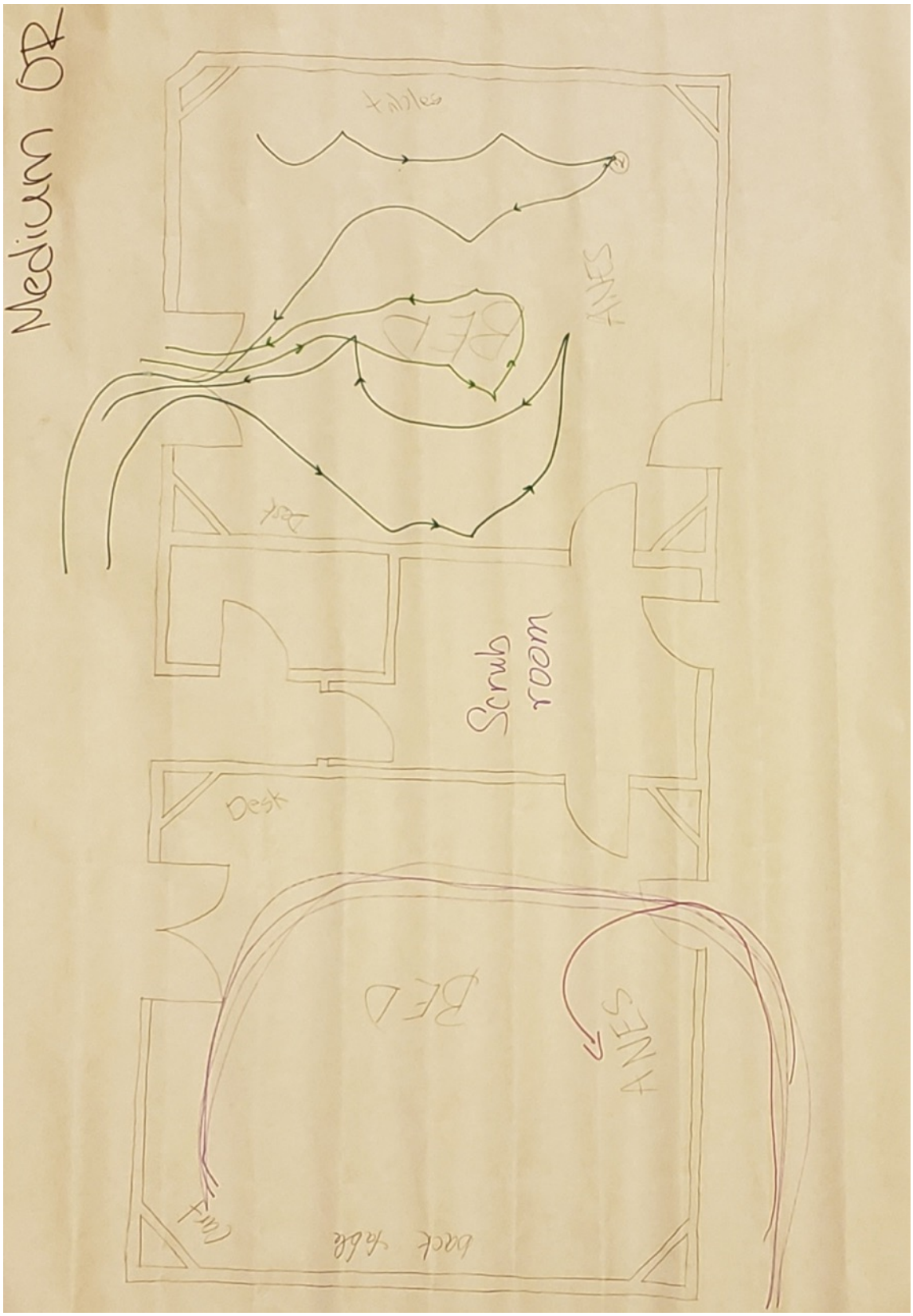


Figure D-2 Operating room spaghetti map

Figure D-3 L2PO/OR spaghetti map

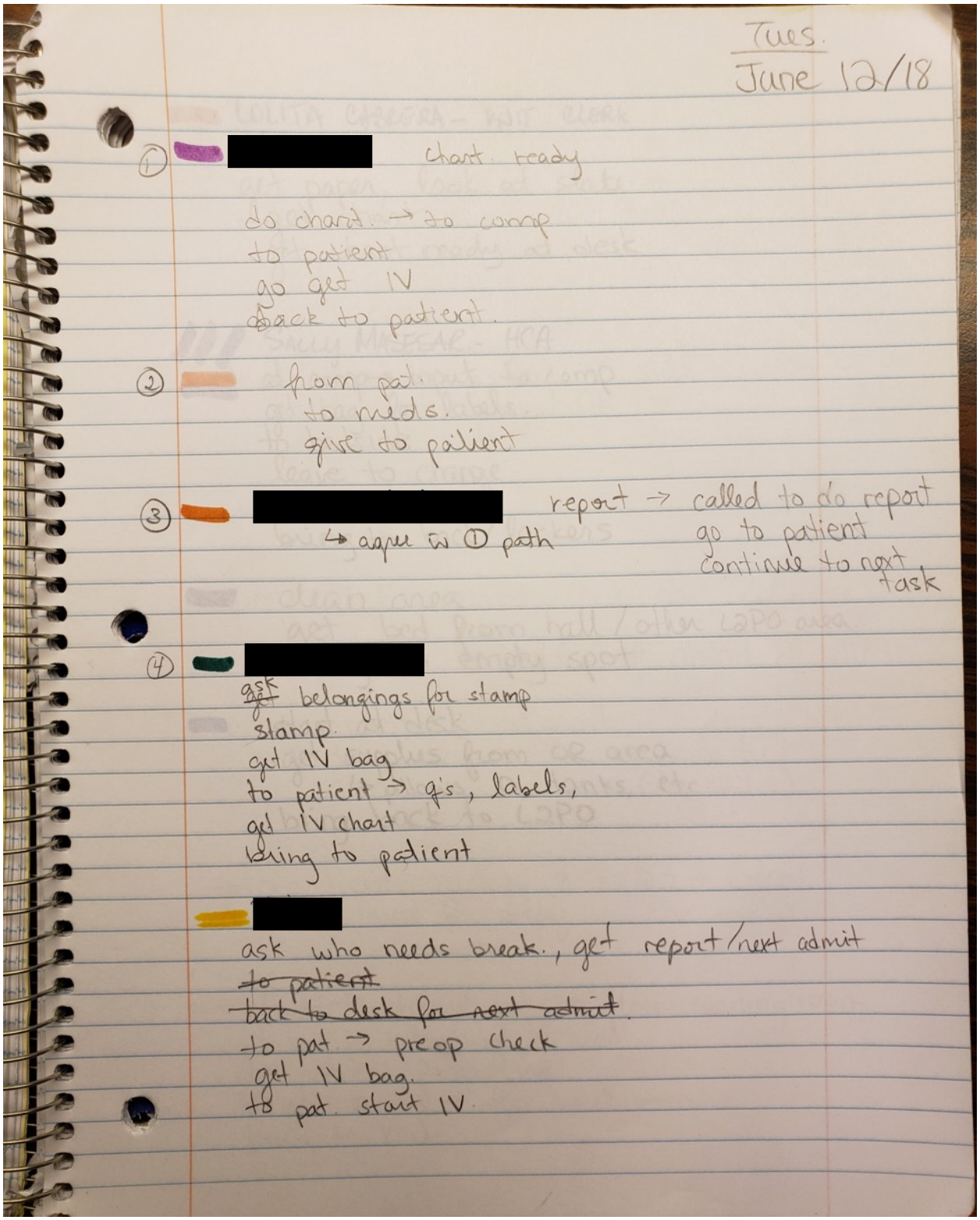


Figure D-4 Spaghetti mapping notes pg. 1

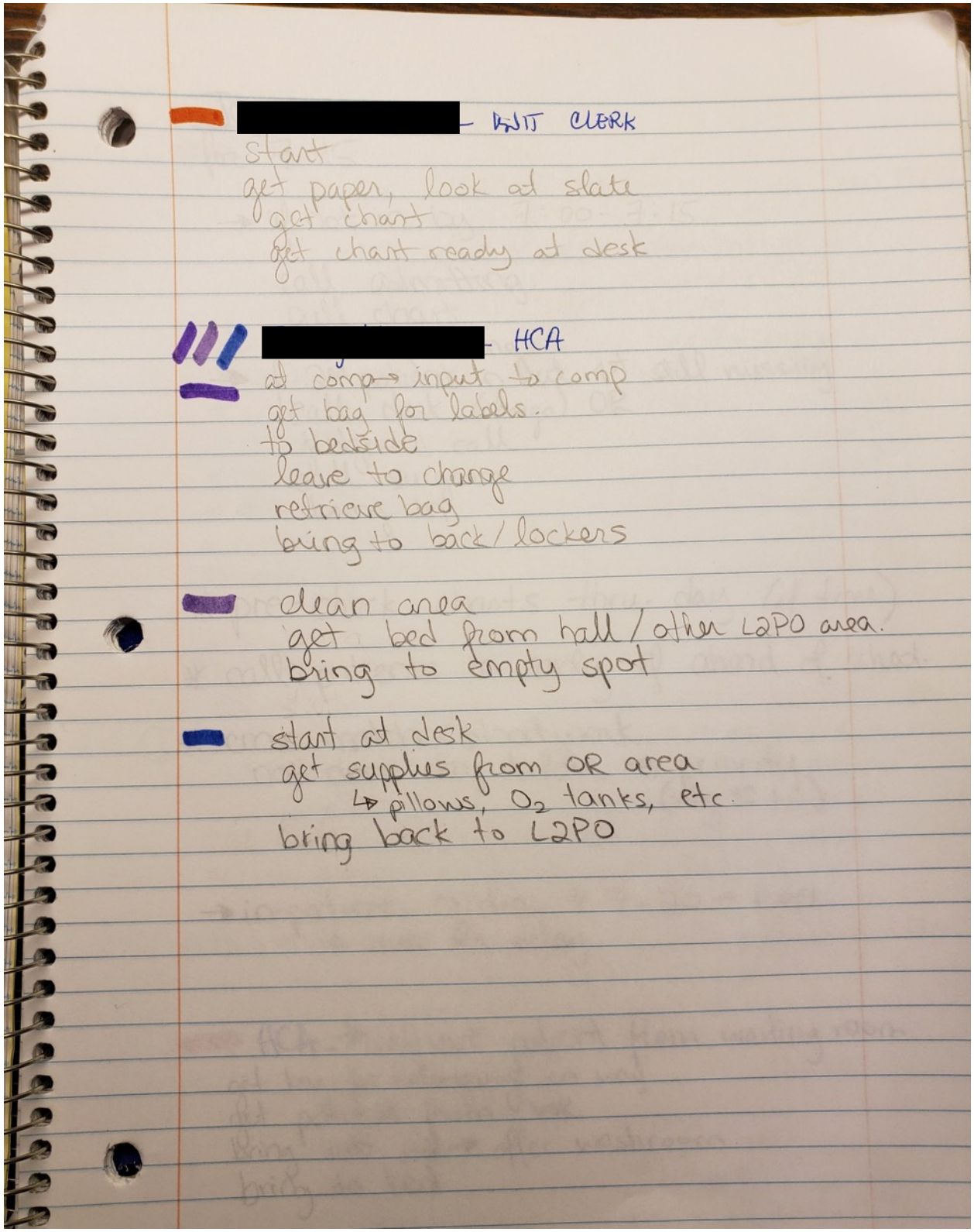


Figure D-5 Spaghetti mapping notes pg. 2

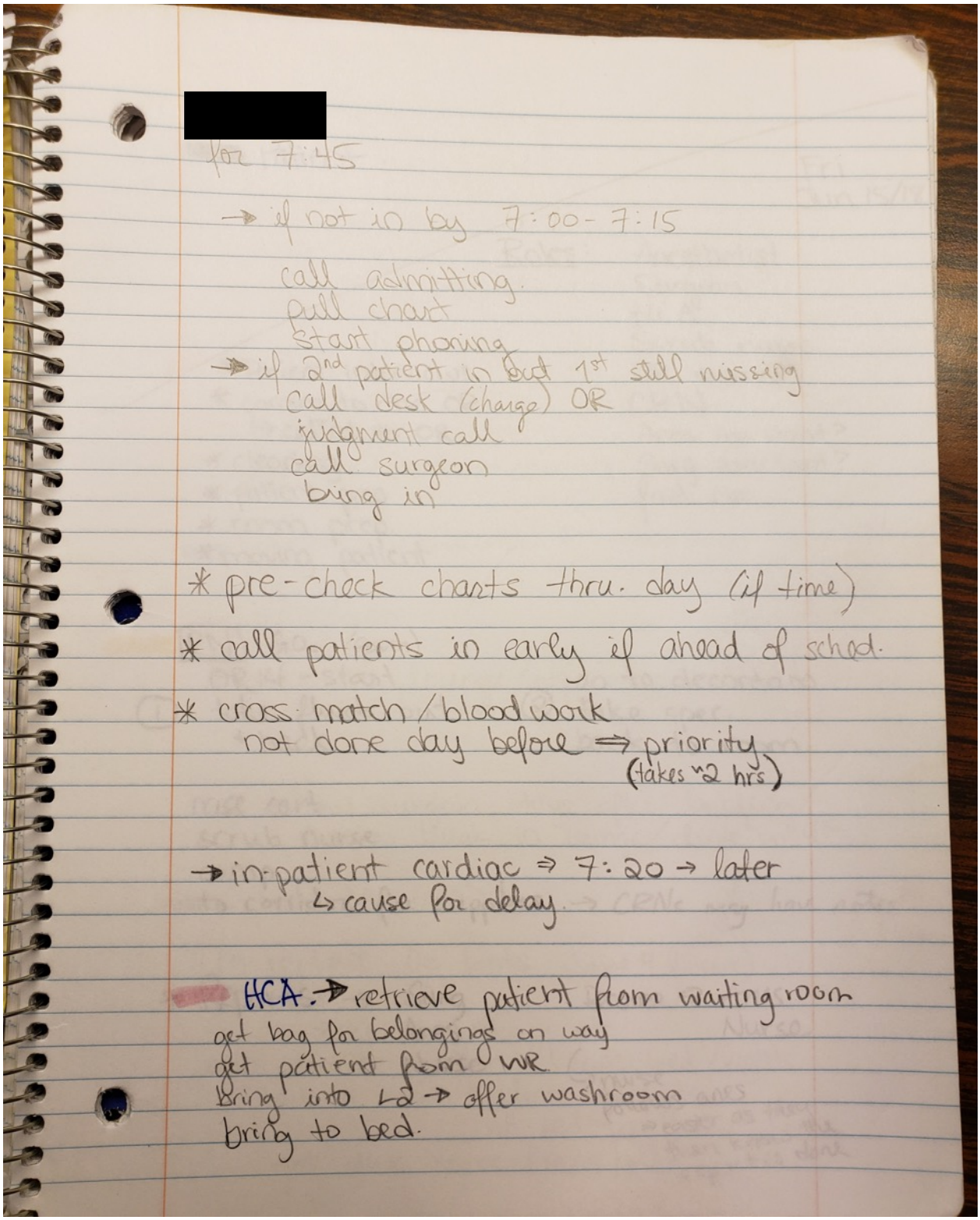


Figure D-6 Spaghetti mapping notes pg. 3

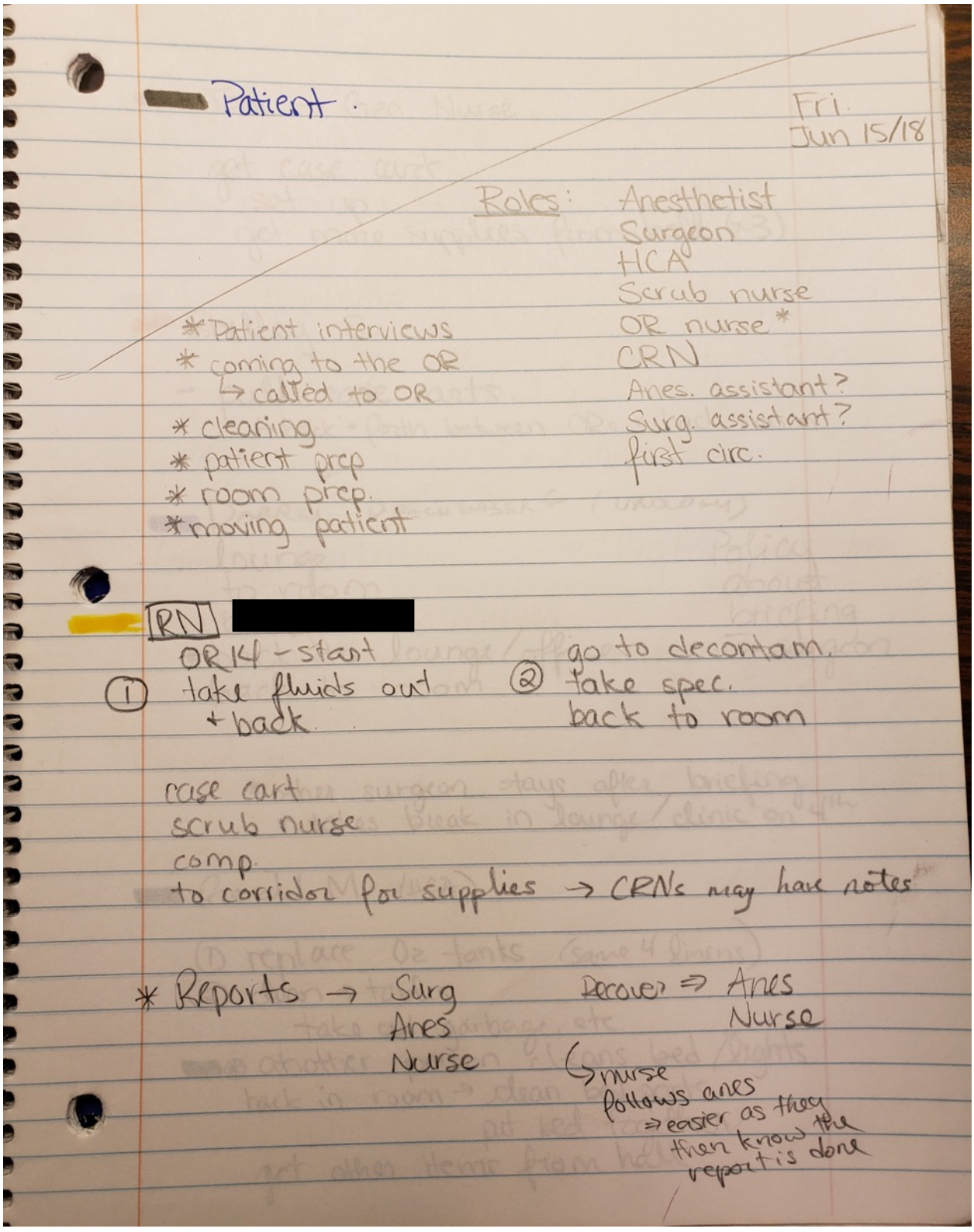


Figure D-7 Spaghetti mapping notes pg. 4

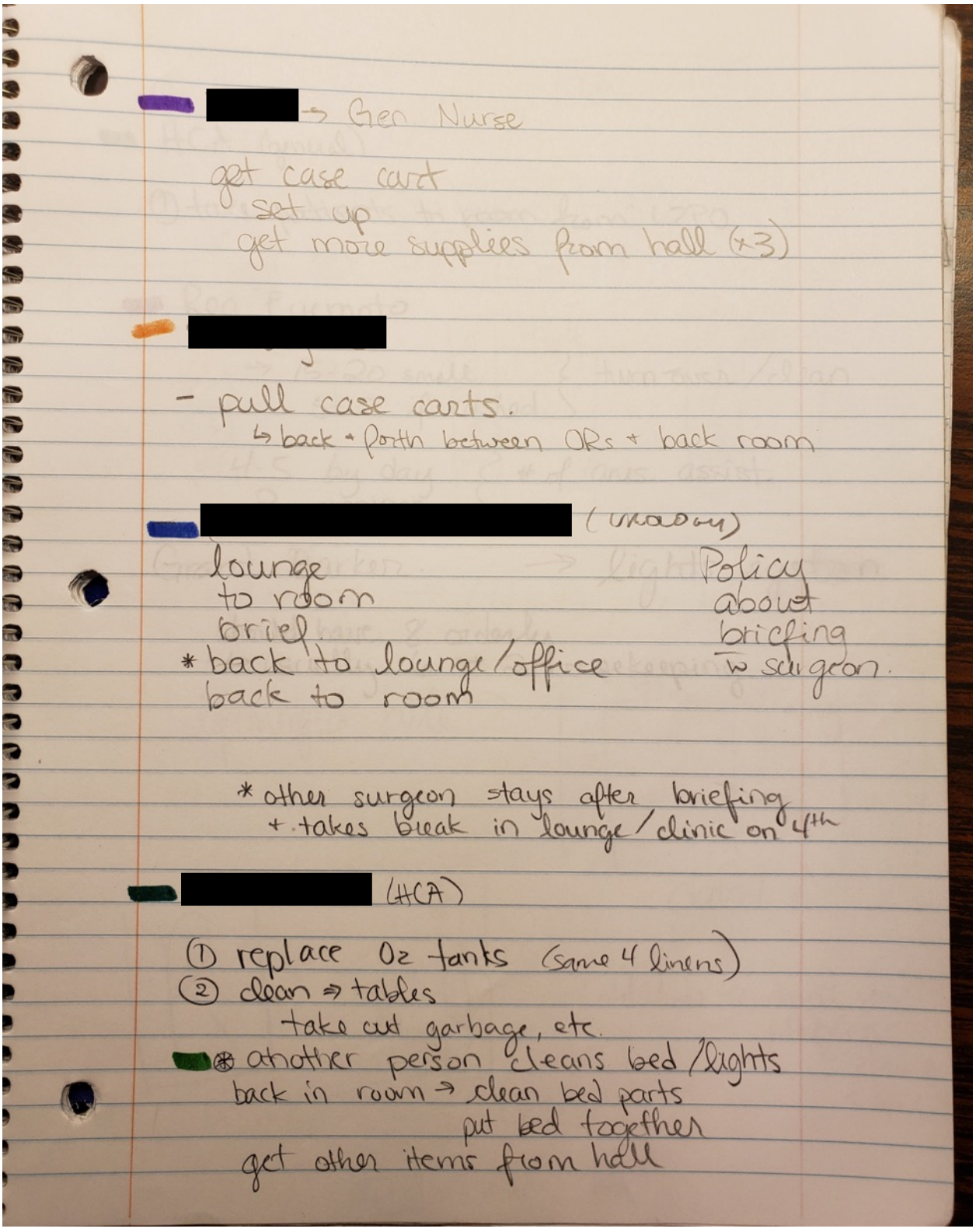


Figure D-8 Spaghetti mapping notes pg. 5

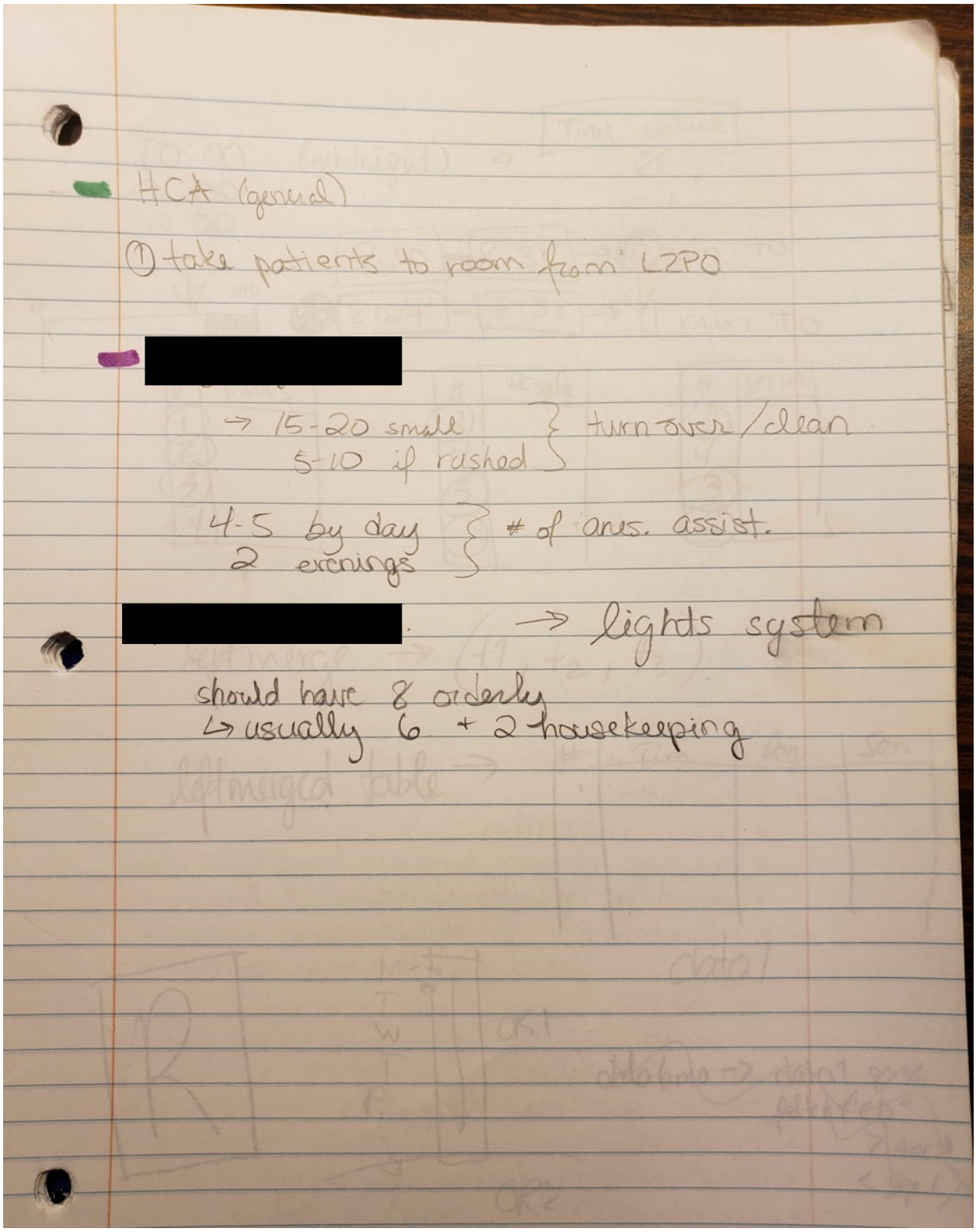


Figure D-9 Spaghetti mapping notes pg. 6

Figure D-10 Sticky note process mapping

APPENDIX E SURGERY INCLUSION SELECTION PROCESS

Three factors were used for choosing surgeries: (1) number of cases, (2) room for improvement, and (3) which surgeons or surgical departments would be affected. Procedures which did not meet the criteria for each factor were eliminated until only six procedures remained.

E.1 Number of Cases

To be included in this project, surgeries must occur frequently enough to have a sufficient proportion occur over a three-month period. To calculate what constitutes a sufficient proportion, Cochran's formula with finite population correction was used (95% confidence level, 50% sample proportion) to calculate the required sample size needed to generate an accurate representation of outcomes for each surgery at three different margins of error (10%, 15%, 20%)

To qualify, the average number of surgeries completed in 3 months had to be greater than the calculated sample size for a 20% margin of error, at minimum. This reduced the possible surgeries from 119 to 16 procedures (Table E-1).

Table E-1 Primary procedures for focus consideration

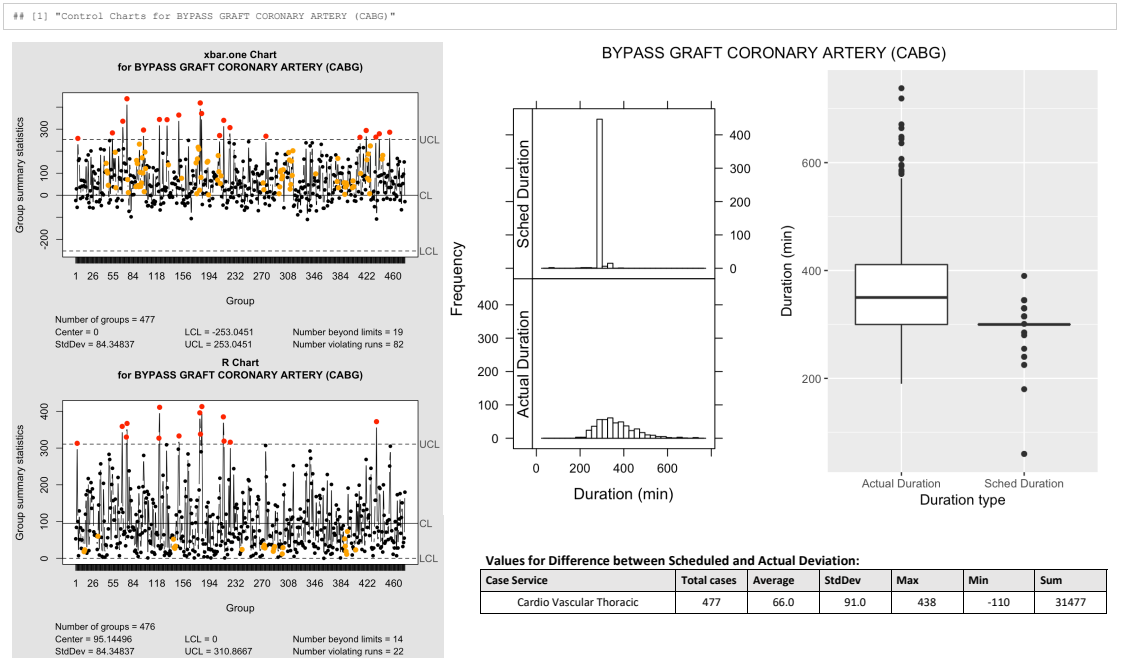
Case service	Primary Procedure	Total cases	20% MOE	15% MOE	10% MOE	3 months	MOE met
cardiac	BYPASS GRAFT CORONARY ARTERY (CABG)	477	23	39	80	119.25	10
otolaryngology	SEPTOPLASTY	210	22	36	66	52.5	15
vascular	FISTULA ARTERIAL VENOUS (AV)	186	21	35	64	46.5	15
OBGYN	HYSTEROSCOPY (DIAGNOSTIC)	182	21	35	63	45.5	15
psychology	STIMULATION BRAIN (ECT)	173	21	34	62	43.25	15
cardiac	REPLACEMENT/REPAIR VALVE AORTIC (AVR)	164	21	34	61	41	15
OBGYN	HYSTERECTOMY ABDOMINAL TOTAL	159	21	34	60	39.75	15
urology	CYSTOSCOPY TRANSURETHRAL RESECTION BLADDER TUMOR CAUTERY MONOPOLAR	151	21	33	59	37.75	15
general	LAPAROSCOPIC CHOLECYSTECTOMY	150	21	33	59	37.5	15
plastic	MAMMOPLASTY REDUCTION	139	21	33	57	34.75	15
head & neck	THYROIDECTOMY	93	19	29	47	23.25	20
otolaryngology	TONSILLECTOMY	83	19	28	45	20.75	20
cardiac	REPLACEMENT/REPAIR VALVE MITRAL (MVR)	82	19	28	44	20.5	20
plastic	ENDOSCOPIC RELEASE CONTRACTURE CARPAL TUNNEL DECOMPRESSION	77	18	28	43	19.25	20
general	LAPAROSCOPIC COLECTOMY HEMI RIGHT	76	18	28	43	19	20
vascular	ENDARTERECTOMY CAROTID	73	18	27	42	18.25	20

E.2 Room for Improvement

Looking only at the 16 procedures remaining, an assessment of which procedures should be included was completed based on baseline data (Figures E1-E16, below). The graphs and data assessed include:

- Mean-Range ($\bar{X}$-R) charts
    - Values: monitor average and control of variability for duration difference
    - Room for improvement, $\bar{X}$: many red/yellow points, high control limits
    - Room for improvement, R: many red/yellow points, centre line further from zero
- Histograms/Boxplots
    - Values: compare frequency distributions and quartile distributions
    - Room for improvement: scheduled vs actual duration distribution very different
- Values on difference between scheduled and actual duration
    - Values: average, standard deviation, maximum, minimum, sum of differences
    - Room for improvement: average further off zero, high standard deviation, large max/min, large positive/negative sum of difference

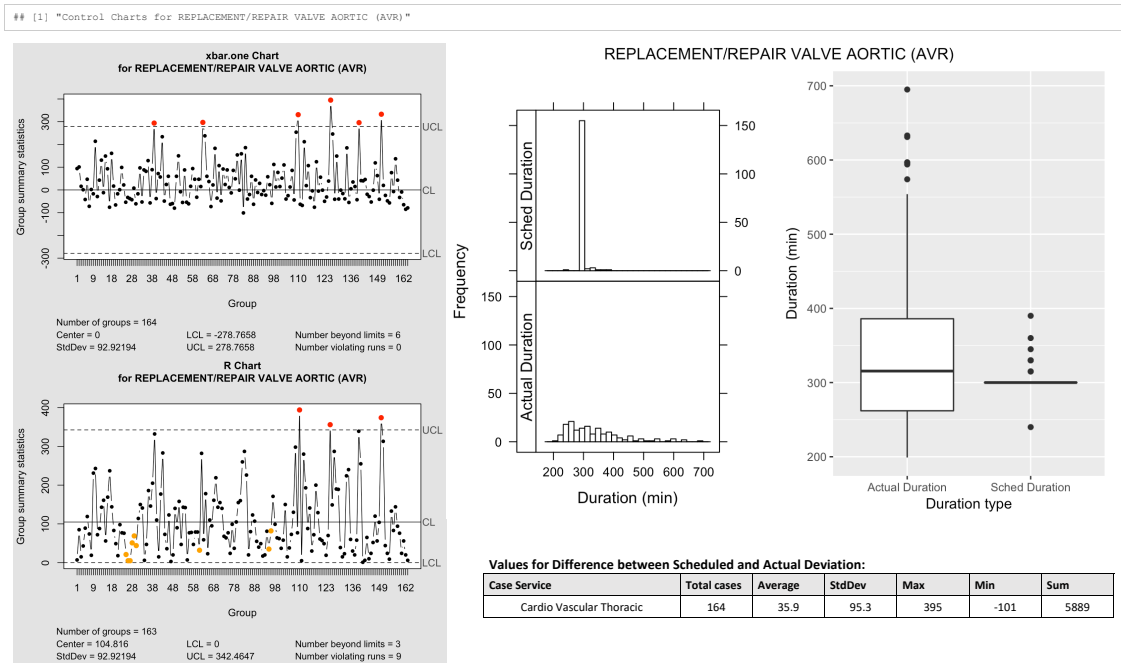
The primary procedures found to have the least room for improvement were tonsillectomy (Figure E-10), laparoscopic colectomy hemi right (Figure E-5), and hysterectomy abdominal total (Figure E-7). The most room for improvement was found within the 3 cardiac surgeries (CABG, Figure E-1; AVR, Figure E-2; MVR, Figure E-3) which consistently run overtime and within the primary procedures hysteroscopy – diagnostic (Figure E-8), fistula AV (Figure E-15), and endarterectomy carotid (Figure E-16) which consistently run undertime.



Values for Difference between Scheduled and Actual Deviation:

Case Service	Total cases	Average	StdDev	Max	Min	Sum
Cardio Vascular Thoracic	477	66.0	91.0	438	-110	31477

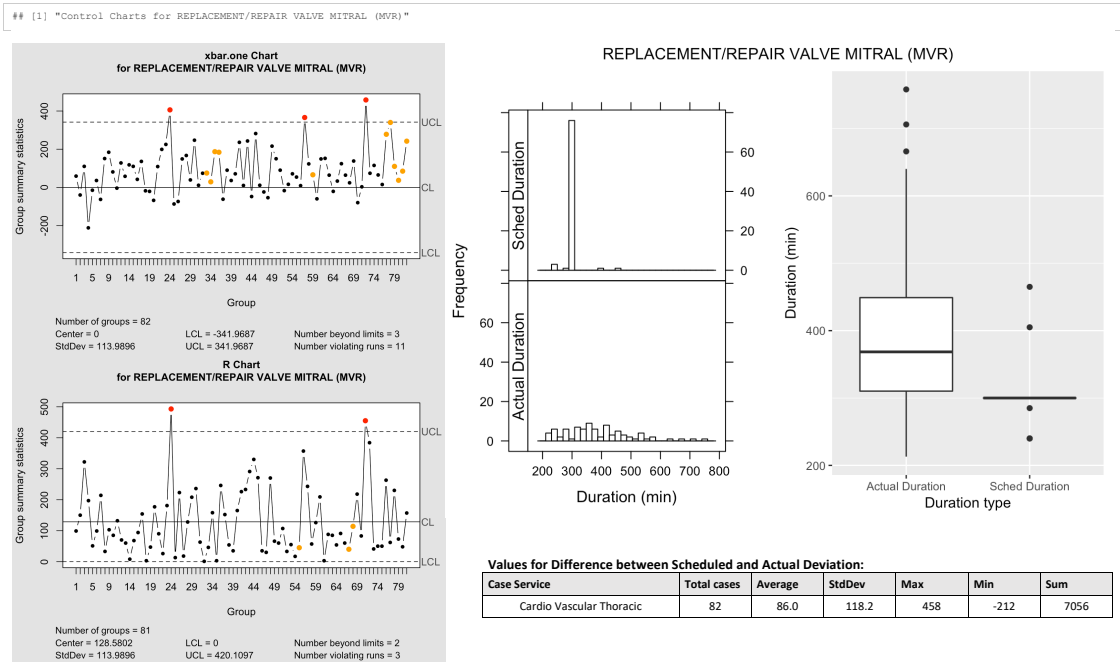
Figure E-1 Cardiac: Bypass graft coronary artery (CABG), baseline data



Values for Difference between Scheduled and Actual Deviation:

Case Service	Total cases	Average	StdDev	Max	Min	Sum
Cardio Vascular Thoracic	164	35.9	95.3	395	-101	5889

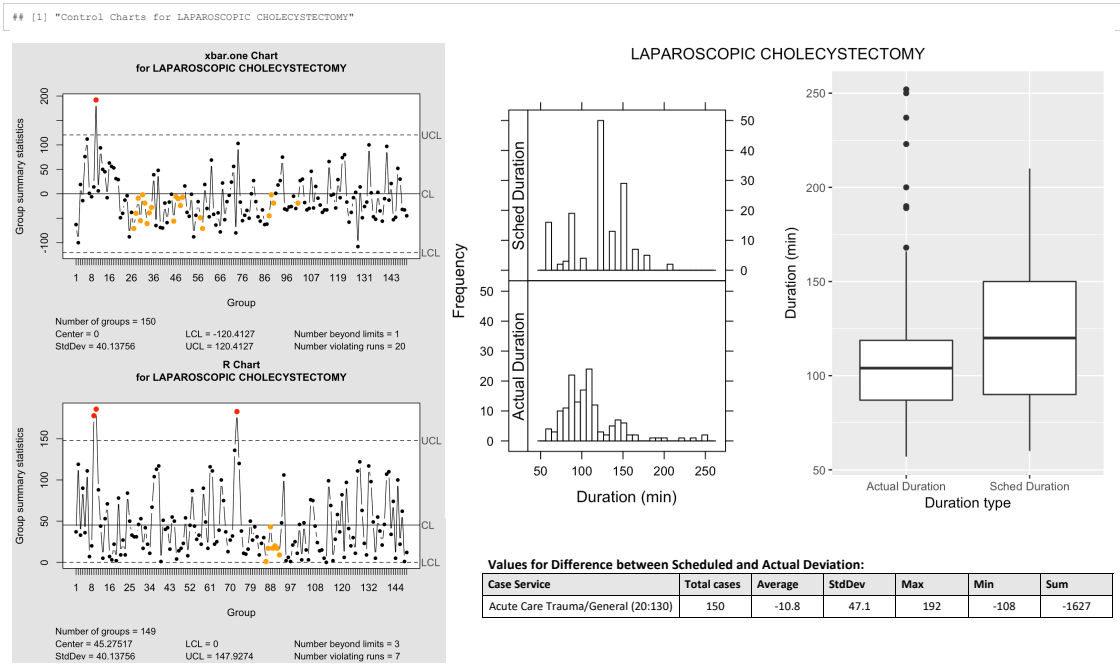
Figure E-2 Cardiac: Replacement/repair valve aortic (AVR), baseline data



Values for Difference between Scheduled and Actual Deviation:

Case Service	Total cases	Average	StdDev	Max	Min	Sum
Cardio Vascular Thoracic	82	86.0	118.2	458	-212	7056

Figure E-3 Cardiac: Replacement/repair valve mitral (MVR), baseline data

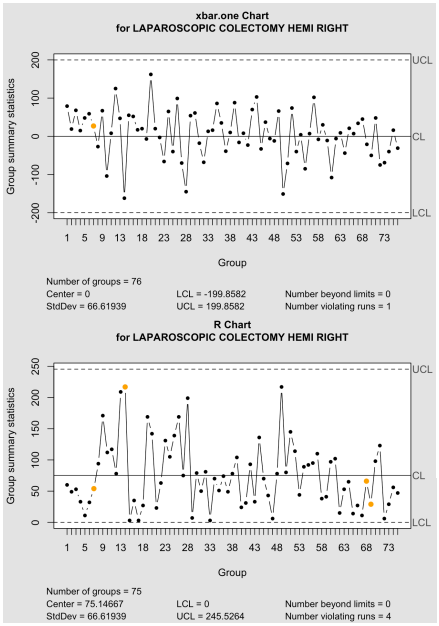


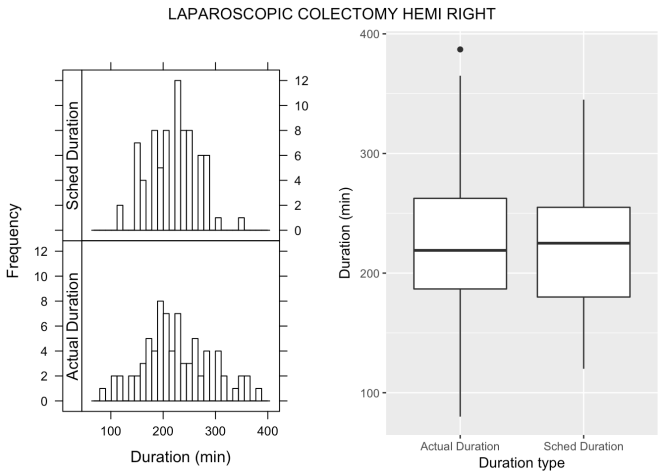
Values for Difference between Scheduled and Actual Deviation:

Case Service	Total cases	Average	StdDev	Max	Min	Sum
Acute Care Trauma/General (20:130)	150	-10.8	47.1	192	-108	-1627

Figure E-4 General/Acute: Laparoscopic cholecystectomy, baseline data

```
## [1] "Control Charts for LAPAROSCOPIC COLECTOMY HEMI RIGHT"
```



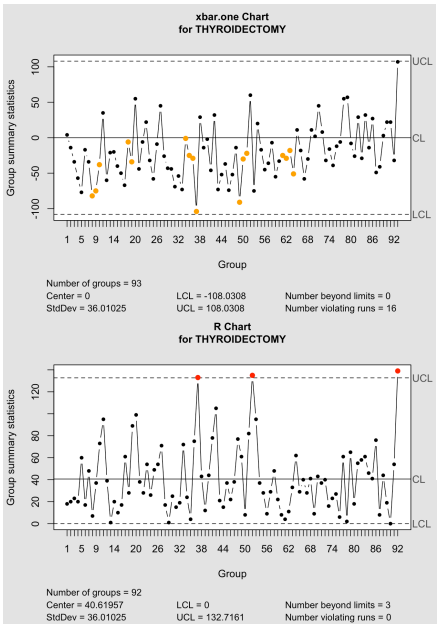


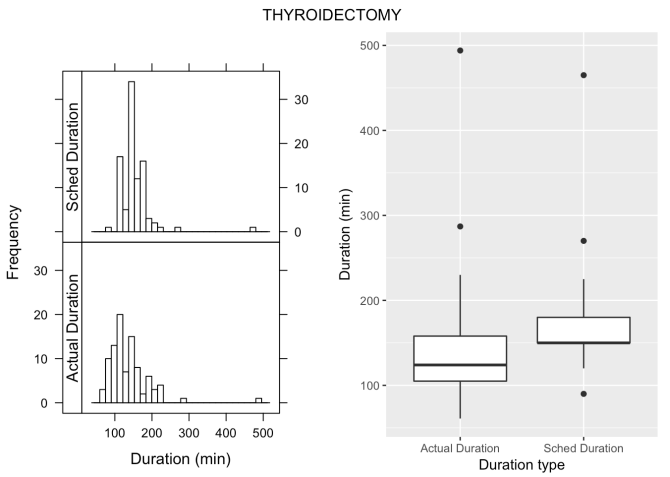
Values for Difference between Scheduled and Actual Deviation:

Case Service	Total cases	Average	StdDev	Max	Min	Sum
General	76	5.9	63.0	162	-162	447

Figure E-5 General/Acute: Laparoscopic colectomy hemi right, baseline data

```
## [1] "Control Charts for THYROIDECTOMY"
```



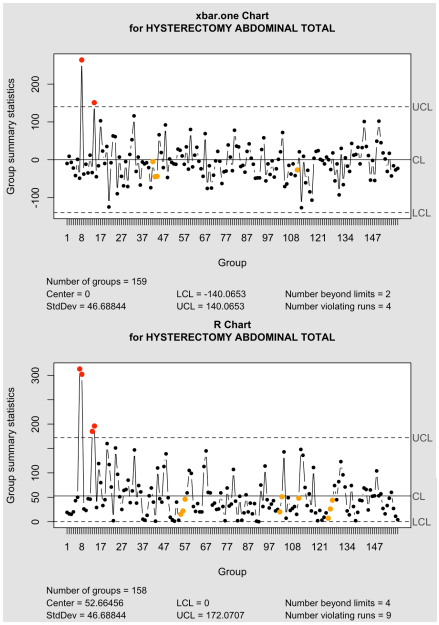


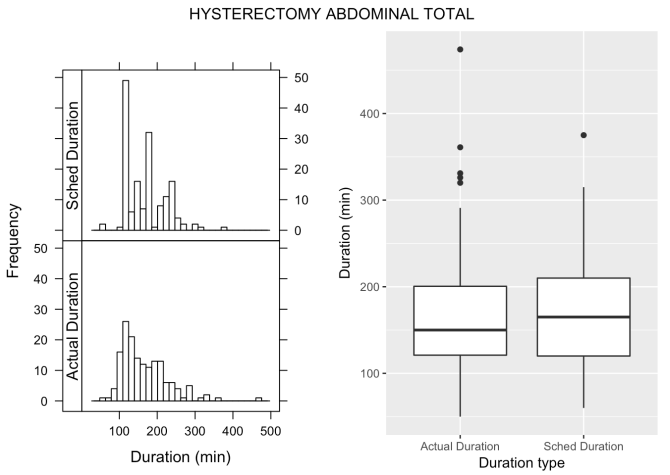
Values for Difference between Scheduled and Actual Deviation:

Case Service	Total cases	Average	StdDev	Max	Min	Sum
Head & Neck/Otolaryngology (54:39)	93	-20.6	38.3	107	-104	-1918

Figure E-6 Head & Neck/Otolaryngology: Thyroidectomy, baseline data

```
## [1] "Control Charts for HYSTERECTOMY ABDOMINAL TOTAL"
```



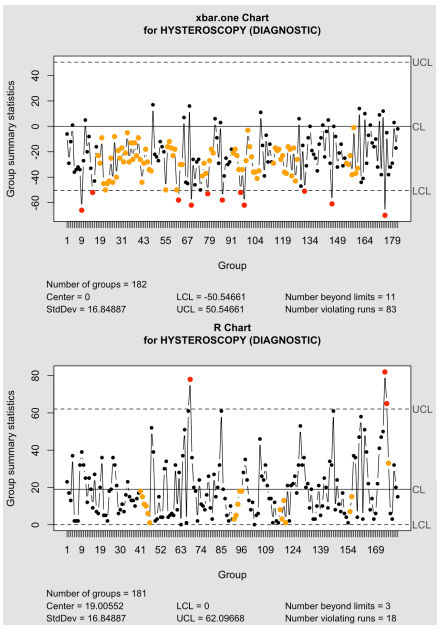


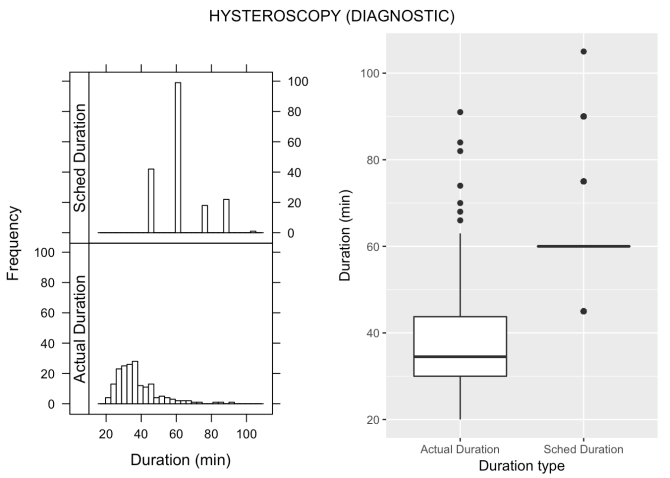
Values for Difference between Scheduled and Actual Deviation:

Case Service	Total cases	Average	StdDev	Max	Min	Sum
Obstetrics and Gynecology	159	-4.9	50.1	264	-127	-785

Figure E-7 OB-GYN: Hysterectomy abdominal total, baseline data

```
## [1] "Control Charts for HYSTEROSCOPY (DIAGNOSTIC)"
```



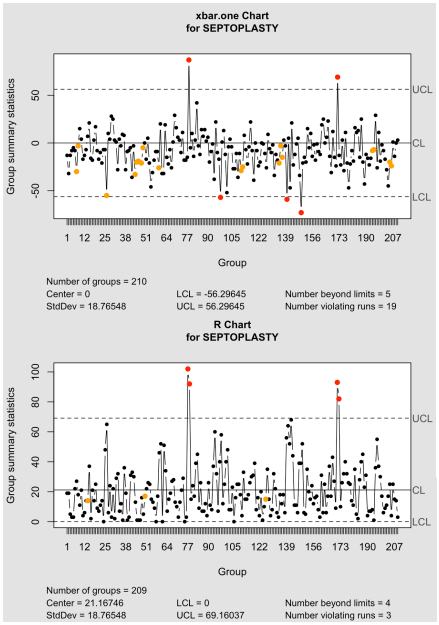


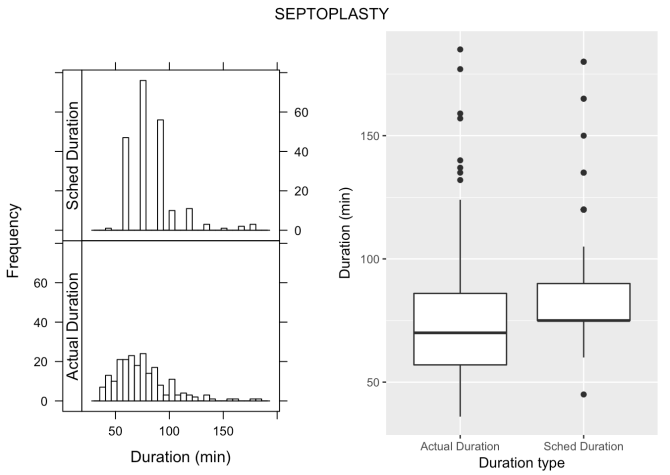
Values for Difference between Scheduled and Actual Deviation:

Case Service	Total cases	Average	StdDev	Max	Min	Sum
Obstetrics and Gynecology	182	-24.4	17.1	17	-70	-4432

Figure E-8 OB-GYN: Hysteroscopy (diagnostic), baseline data

```
## [1] "Control Charts for SEPTOPLASTY"
```



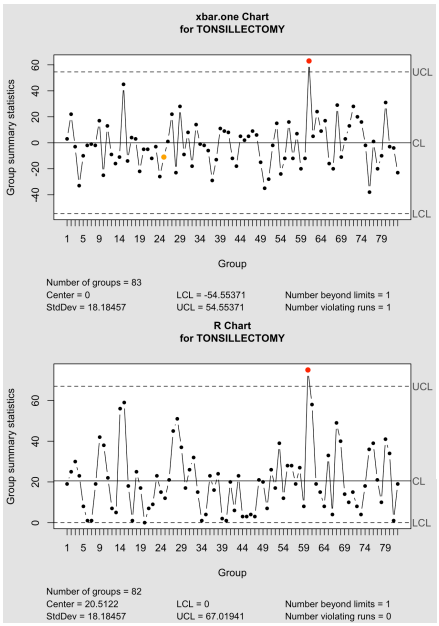


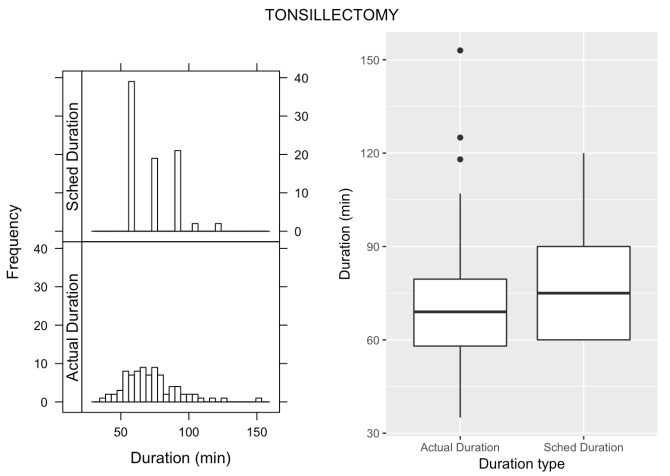
Values for Difference between Scheduled and Actual Deviation:

Case Service	Total cases	Average	StdDev	Max	Min	Sum
Otolaryngology	210	-8.4	20.1	87	-73	-1762

Figure E-9 Otolaryngology: Septoplasty, baseline data

```
## [1] "Control Charts for TONSILLECTOMY"
```

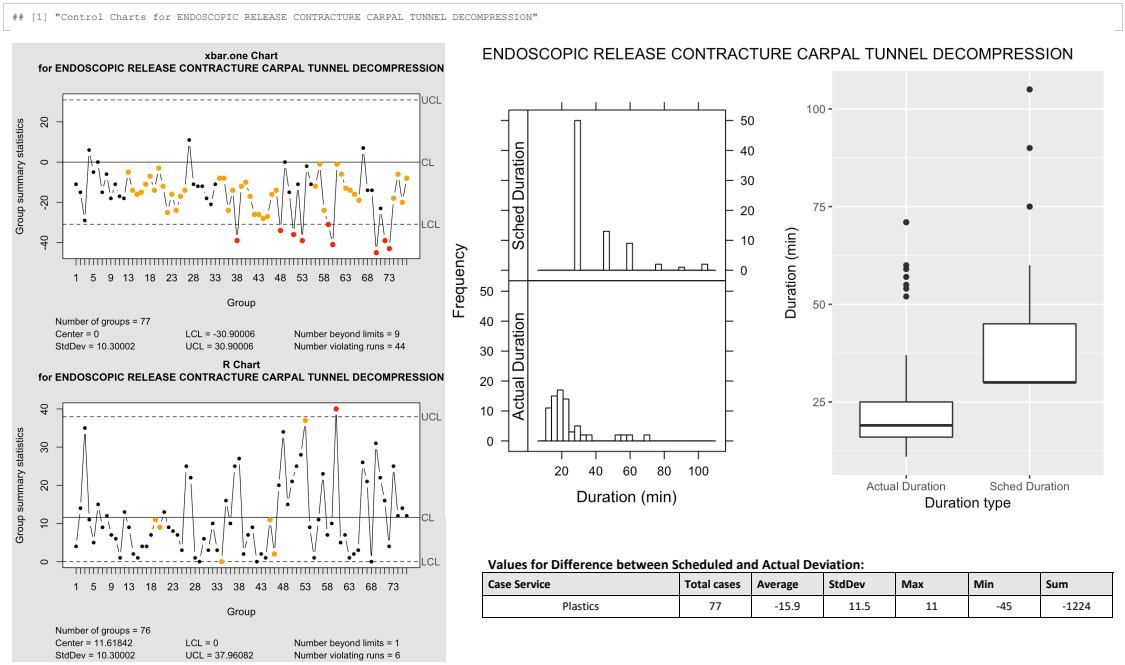




Values for Difference between Scheduled and Actual Deviation:

Case Service	Total cases	Average	StdDev	Max	Min	Sum
Otolaryngology	83	-1.4	18.2	63	-38	-116

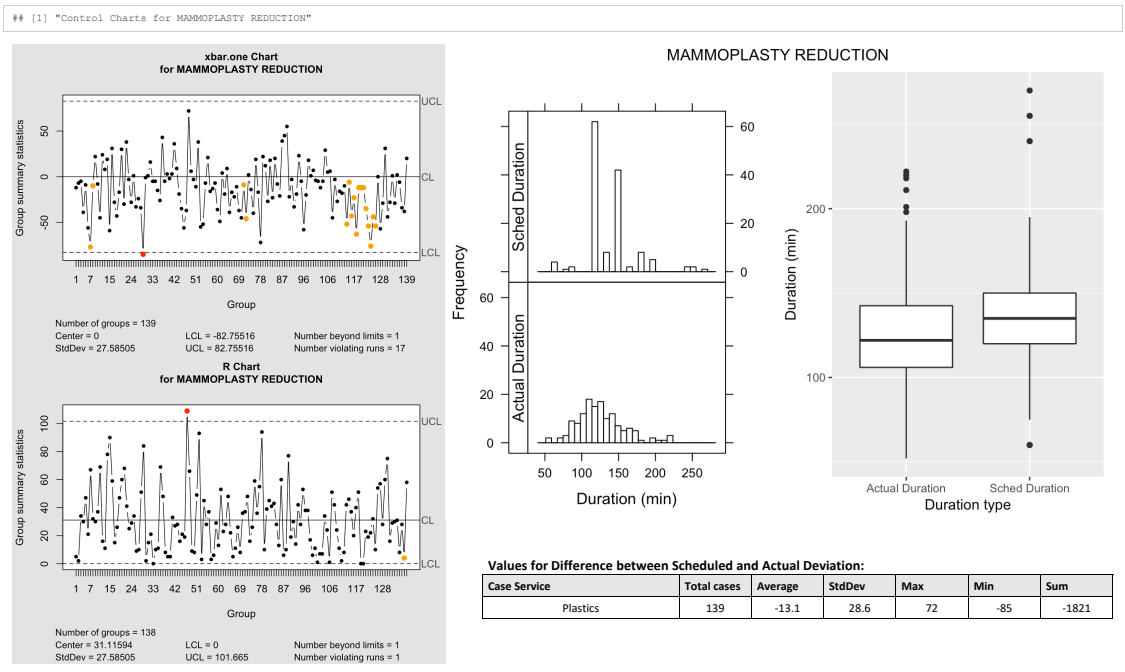
Figure E-10 Otolaryngology: Tonsillectomy, baseline data



Values for Difference between Scheduled and Actual Deviation:

Case Service	Total cases	Average	StdDev	Max	Min	Sum
Plastics	77	-15.9	11.5	11	-45	-1224

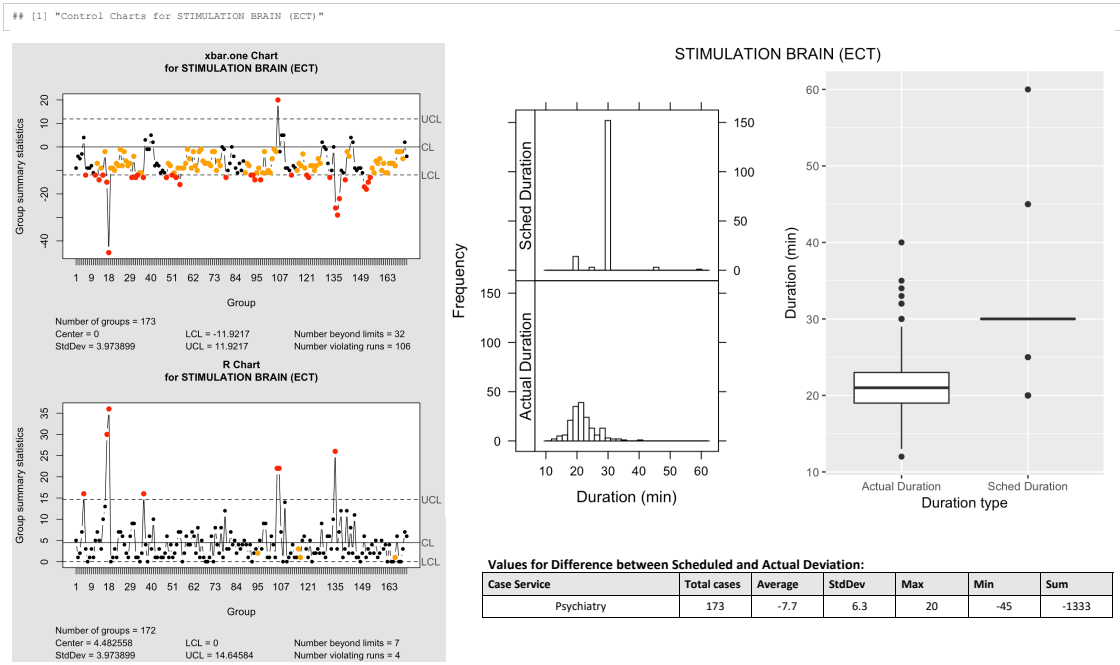
Figure E-11 Plastics: Endoscopic release contracture carpal tunnel decomp, baseline data



Values for Difference between Scheduled and Actual Deviation:

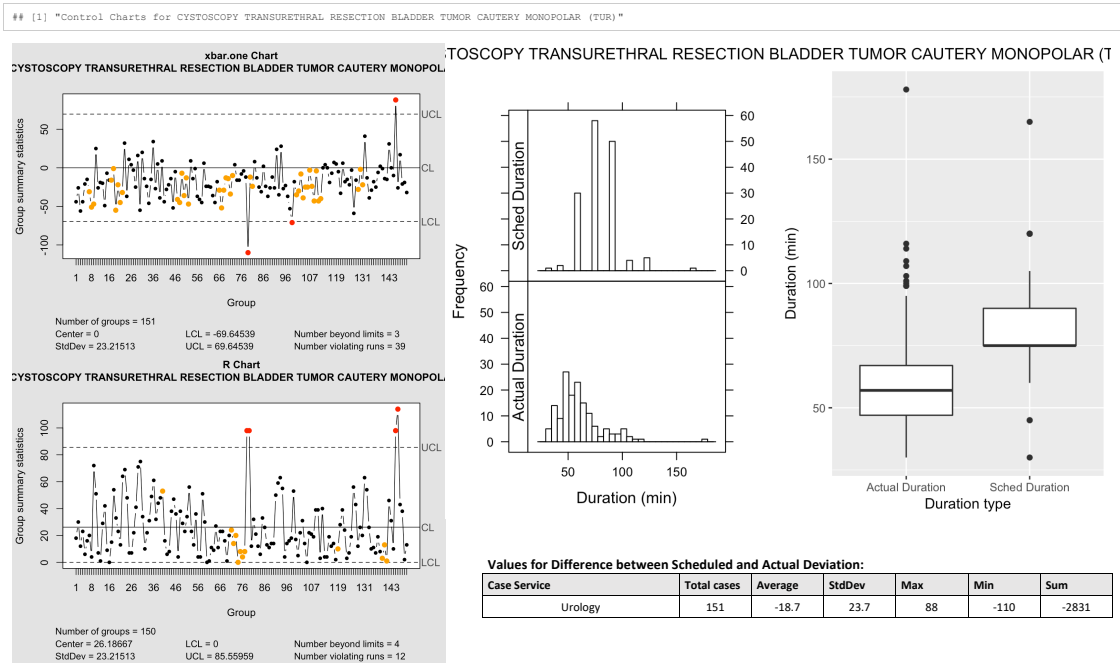
Case Service	Total cases	Average	StdDev	Max	Min	Sum
Plastics	139	-13.1	28.6	72	-85	-1821

Figure E-12 Plastics: Mammoplasty reduction, baseline data



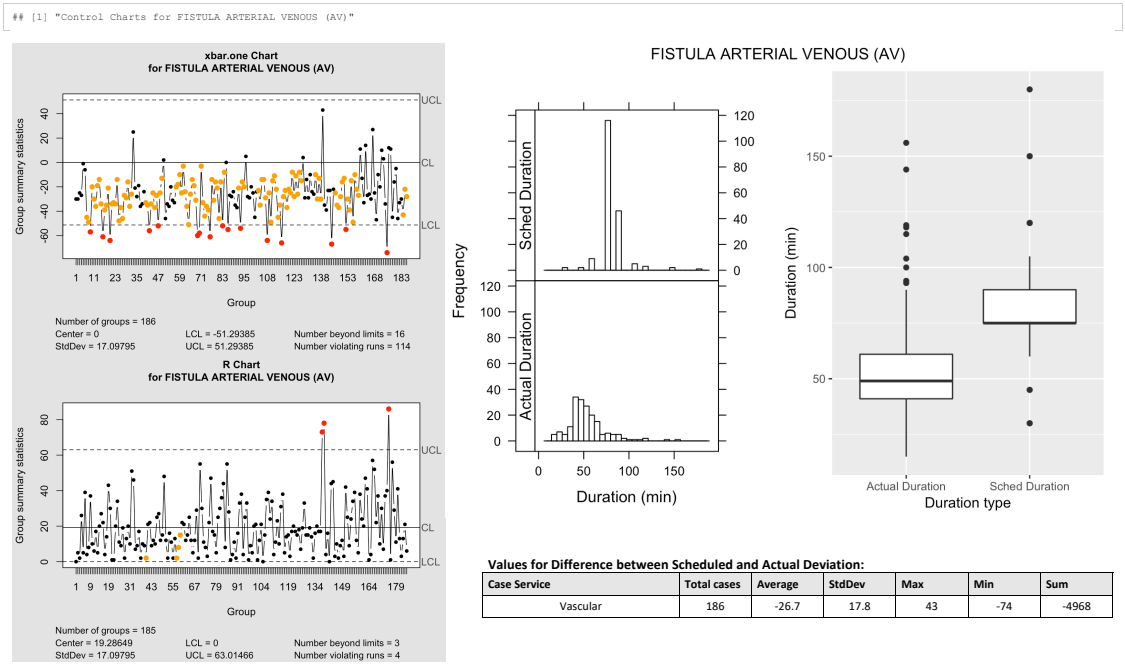
Case Service	Total cases	Average	StdDev	Max	Min	Sum
Psychiatry	173	-7.7	6.3	20	-45	-1333

Figure E-13 Psychiatry: Stimulation brain (ECT), baseline data



Case Service	Total cases	Average	StdDev	Max	Min	Sum
Urology	151	-18.7	23.7	88	-110	-2831

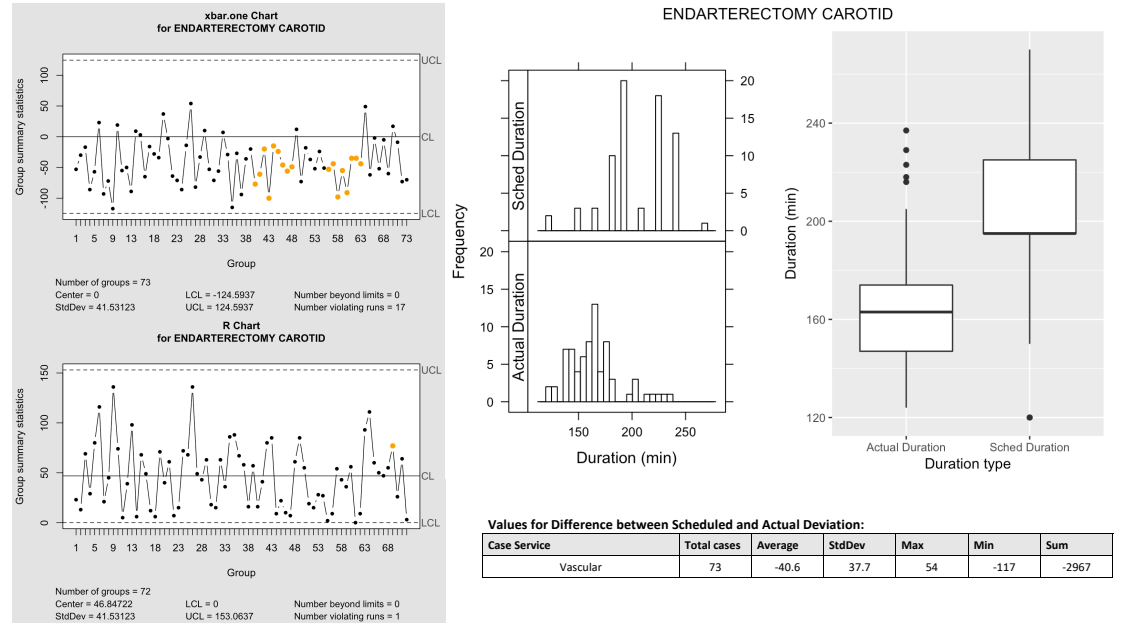
Figure E-14 Urology: Cysto. transur. resection bladder tumor cautery mono, baseline data



Values for Difference between Scheduled and Actual Deviation:

Case Service	Total cases	Average	StdDev	Max	Min	Sum
Vascular	186	-26.7	17.8	43	-74	-4968

Figure E-15 Vascular: Fistula arterial venous (AV), baseline data



Values for Difference between Scheduled and Actual Deviation:

Case Service	Total cases	Average	StdDev	Max	Min	Sum
Vascular	73	-40.6	37.7	54	-117	-2967

Figure E-16 Vascular: Endarterectomy carotid, baseline data

E.3 Surgeons Affected

Based on the guidance and recommendations of the executive team, the 16 procedures were further narrowed down based on the surgeons and teams that would be affected. We chose to eliminate all surgeries relating to the case services of plastics, otolaryngology, psychiatry, head & neck, urology, and general surgery. The remaining case services are cardiac, obstetrics and gynecology (OB-GYN), and vascular. Within the vascular procedures, we also chose to eliminate the fistula arterial venous procedure. This left us with only six procedures remaining.

E.4 Chosen Surgeries for Inclusion

The six procedures and related case services which were chosen to be included for testing and improvement are:

- Cardiac: Bypass graft coronary artery (CABG) (Figure E-1)
- Cardiac: Replacement/repair valve aortic (AVR) (Figure E-2)
- Cardiac: Replacement/repair valve mitral (MVR) (Figure E-3)
- OB-GYN: Hysterectomy abdominal total (HAT) (Figure E-7)
- OB-GYN: Hysteroscopy – diagnostic (HYSD) (Figure E-8)
- Vascular: Endarterectomy carotid (END) (Figure E-16)

APPENDIX F **SHORT-TERM DATA SOLUTION**

As a short-term attempt to improve case duration estimate accuracy, historical data were compiled and summarized for each surgeon in St. Boniface Hospital. The data provided to the executive team were grouped by case service, main surgeon, primary procedure description, and number of secondary procedures, for which the number of outputs provided is shown in Table F-1. Values for the number of times a surgery was performed by a specific surgeon were calculated, along with time data. The data were also filtered for surgeries which were performed a minimum of ten times by a surgeon. The time data calculated included averages and trimmed averages (top and bottom 10% removed) of the estimated duration, actual procedure duration, actual case duration, and error in minutes between the estimate and actual case duration.

Table F-1 Surgeon-Primary procedure data

Grouped by	Filter	Number of Outputs
Primary Procedure Description Main Surgeon	NA	1898
Primary Procedure Description Main Surgeon	10+ occurrences	310
Primary Procedure Description Main Surgeon Number of Secondary Procedures	NA	3151
Primary Procedure Description Main Surgeon Number of Secondary Procedures	10+ occurrences	261

A summary of the error rate in minutes was also provided to the executive team for each surgeon in each case service, regardless of procedure. The data were grouped based on the case classification (overtime cases, undertime cases, cases which were on-time/within 15 minutes, and total cases) and the number of cases associated with each classification were calculated, as well as the average error of the actual case duration from the estimate in minutes. The data for each surgeon in each case service were organized as shown in Table F-2.

Table F-2 Surgeon error data layout

Case Service	Main Surgeon	Classification	# of Cases	Avg Error
		Over		
		Under		
		Within 15 min		
		Total		

The executive team distributed the data as they saw fit. This was completed outside of the scope of this project, so the outcomes of this implementation of data were not observed or recorded. Due to the scope of this thesis and related Lean Six Sigma project being limited to small scale implementation, it was decided that this could have a beneficial effect on case duration estimate accuracy while more research and testing is still needed toward implementing predictive modeling or historical average methods.

APPENDIX G **PREDICTIVE MODELS**

First, research was done on the past and current methods for surgical duration estimates. Once the research was completed, the process of creating the predictive models was undertaken. The process used to create the predictive models consisted of four parts: editing for errors and outliers, data exploration, model generation, and model testing. Only the data for the six chosen procedures were used in this process. The core coding used to create the models is also provided.

G.1 Surgical Duration Estimates Research

G.1.1 Predictive Models for Surgery Duration

Previous research on surgical duration estimates has shown that surgeon estimates can be improved through the use of historical data, such as with moving or trimmed averages [21] [22] [23] [24] [25]. Depending on the variables included in the historical averages, the estimates may be more accurate; calculating based on surgeon and case complexity can improve the estimate [25], as can adding the standard deviation to the mean to predict possible over-booking [23]. However, the historical averages can be improved even further through the use of predictive modeling and machine learning methods. There are many types of machine learning and predictive modeling methods that have been tested, such as ANOVA [26], Bayesian models [27] [28] [29], Mixture Density Network [29], Neural Networks [30] [31], Support Vector Regression [32], Multivariate Adaptive Regression [21], Lognormal models [33] [34], and Random Forests [21] [35]. Linear models (regression [21] [22] [36] [37] [38] [39] [40] [41], generalized [35], mixed [42]) are some of the more commonly tested approaches to modeling surgical durations.

For this project, I've chosen to look further into three different types of models: simple moving average (SMA), linear mixed effects (LME), and random forest (RF).

A SMA model is one of the easiest and simplest ways to predict future surgery durations. It only requires that a certain number of cases be completed in order to use the historical data to create an estimate. A drawback of this method is that it does not account for the effects of any other variables and cannot be used for procedures which occur rarely.

A linear regression model is more complex than a SMA model as it includes more input variables, however this also may allow it to provide a more accurate prediction. The LME model is a type of linear regression model. While a standard linear model has only fixed effects, the LME model also includes random effects which is particularly useful when dealing with non-independence or hierarchical data [43]. Interpreting fixed and random effects can vary depending on viewpoint, so there is not a definitively correct or incorrect assignment [44] [45]. However, there are some general guidelines that should be followed. Firstly, continuous variables are always fixed while categorical variables may be either fixed or random [45] [46]. For categorical variables to be considered a fixed effect, the data should be gathered from all levels of interest or be taken from specific levels of interest as set beforehand [44] [46] [47]. For example, from this data set the day of the week could be considered as a fixed effect because procedures from all weekdays are included and weekends were not included on purpose because elective surgeries are not performed on weekends. Alternatively, random effects may either be used to control for the effect of grouping factors [45] or have many factor levels for which all levels are of interest but not all are included in the data [45] [46]. In addition, random factors may be considered as nested if the factor levels seen for a variable are specific to a certain group and have no cross-over to other groups [45]. For example, from this data set the case service is a random effect because the procedures are grouped within the factor levels and surgeons are a nested variable within case service because they only work within one case service (no cardiac surgeons are

working on OB/GYN procedures or vice-versa); also, not every surgeon which could perform the surgery may be represented in the data. Essentially, for a fixed effect we are interested in the specific differences between all levels, while for random effects only the general impact of the levels is required. One more requirement for random effects is that they should have at least 5 factor levels or else should be considered as fixed [45] [46]. Random effects estimate variance between factor levels and if there are less than five levels then the estimates will be imprecise and unlikely to benefit the model.

The RF model is an ensemble learning method which uses bootstrap aggregation (or bagging) in order to utilize a "forest" of decision trees running in parallel [48]. A decision tree splits at multiple points based on the input variables until the reaching the output or the "leaves" on the tree. In a random forest, each tree is built using a different random sample of rows from the training data and splits based on different random selections of the input variables available. The output value is then calculated as the mode from all the trees if a categorical value, while for a regression the output is calculated as the mean of the results obtained from each tree. This improves on linear regression methods because it can better account for nonlinear relations between input and output variables, particularly for categorical variables. One drawback of this method is in extrapolation—it fails if a new factor for a categorical variable is used which was not in the training data and it cannot identify trends which extend past the training data [49]. For new factor levels, a possible solution is to create a factor level which combines uncommon or rare levels so that a new factor can be grouped in with other new or rare values, though it may result in a slightly less accurate output. As for extrapolation of trends, it is not an issue in this case as the duration estimates are expected to fall within the range of the data used.

G.1.2 Surgery Duration Variables

Another area to consider when modeling is which explanatory variables to include. The most influential and commonly included variables were case service/specialty, procedure type, and surgeon, of which some combination of these variables was included in all surgical modeling articles referenced [21] - [42] and all are available in the current data. The surgeon's estimate was also included in multiple models [29] [33] [38] [42] and was found to be highly influential to the prediction, so will also be considered for inclusion. Patient information was another commonly included variable, such as age, sex, and patient status [35] [38] [39] [40] [42], though they were not always found to be significant [38] [42]. These three patient variables are also available in the current data and will be assessed for inclusion in the model. Other variables that are available in the current data set and found in literature include anesthesia information (ASA score, anesthesiologist, anesthesia type) [29] [34] [38], number of secondary procedures [33] [40] [42], and OR information (room number [35], day of the week [29] [39]).

There are also variables that were used in references which are not available in the current data set. Patient information that was sometimes used but not available for this project includes body mass index (BMI) and medical history [40] [42]. Additional information that could be relevant to the surgery duration but which was not provided for this project includes the familiarity of the surgical staff with each other [50] [51] and—as this project is occurring in a teaching hospital—whether a resident or young surgeon (< age 30) will be participating in the surgery, as additional time may be needed to accommodate teaching during the surgery [42]. These variables may be relevant to the duration, therefore additional research is needed to determine impact and whether effort should be made to include these variables in future record-keeping. Daily workload is another variable that has been used [37] [39] and could be

extrapolated from the data. However, the number of cases completed in a day is dependent on how long the scheduled surgeries take, thus the variable is actually dependent on the outcome. A summary of the possible variables to include is given in Table G-1.

Table G-1 Surgical duration variables (~~italics~~ = variable not available in current data)

Patient Variables	Surgery Variables	Anesthesia Variables	Room/Staff Variables
Age	Case service	ASA score	Room number
Sex	Procedure	Anesthesia type	Day of the week
Patient status	# of procedures	Anesthesiologist	Primary surgeon
~~*BMI*~~	Surgeon estimate		~~*Staff familiarity*~~
~~*Medical history*~~	~~*Resident included*~~		~~*Daily cases*~~

G.2 Data Editing

Before analysis, the data needed to be checked for errors and outliers. Only outliers which are caused by special causes or errors were eliminated—not all outliers—since outliers due to common causes are still an expected part of the data distribution. However, extreme outliers due to special causes could skew the model, so should be removed. Any durations less than zero were eliminated as they would be caused due to errors in data collection; if the times between the case elements are not properly recorded, it may result in a duration being in the negative. This resulted in 17 cases being removed. The elimination of these cases and how many were eliminated from each procedure is shown in Table C-2, along with the previous data filtering steps.

No other errors were found in the data. For the 1120 remaining cases, all individuals were between the ages of 20 and 91. The case start times were all within range; though some cases began after end-of-day and went even further over-time, the duration values associated with these cases did not show any error. Late or over-time cases were not considered as errors.

Outliers were identified as cases with both procedure duration and case duration being above an upper control limit (UCL), which is calculated as the mean value plus three standard

deviations. Both duration values had to be above the associated UCL in order for the surgery to be removed. By only removing the cases which are outliers in both areas it ensures that the actual procedure time is significantly higher than normal, and that time is not recovered by being faster in other areas (preparation, anesthesia, wrap-up). This resulted in a total of 15 cases being removed, less than 2% of the total number of cases for each procedure (Table G-2). $\bar{X}$ plots were used to visualize and identify these outliers (Figure G-1).

Table G-2 Procedure/Case duration outliers

Procedure	Procedure duration (UCL)		Case duration (UCL)		Both outliers (% total)	
CABG	9	(> 507.6)	7	(> 600.9)	7	(1.5%)
AVR	5	(> 506.2)	3	(> 599.2)	3	(1.9%)
MVR	0	(> 631.3)	1	(> 735.8)	0	(0.0%)
END	1	(> 150.1)	2	(> 226.0)	0	(0.0%)
HAT	3	(> 243.9)	2	(> 337.3)	2	(1.3%)
HYSD	4	(> 33.6)	5	(> 69.2)	3	(1.7%)

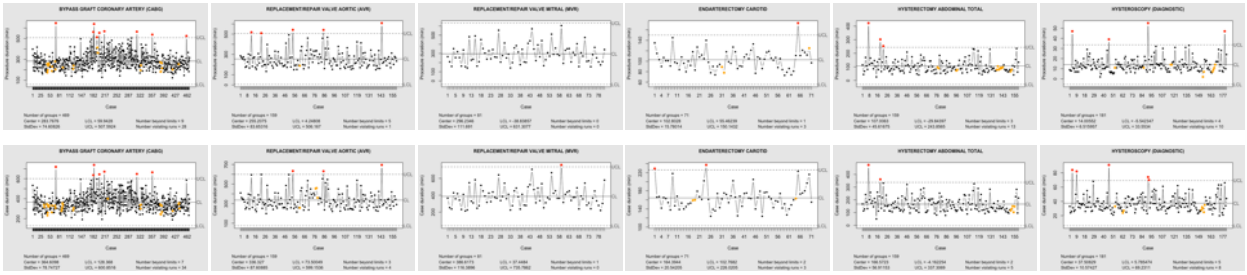


Figure G-1 $\bar{X}$ plots for procedure and case durations to identify common outliers

Anesthesia durations were found to have outliers both above the UCL and below the lower control limit (LCL). These outliers were identified based on whether they were past the control limit and also more than a full standard deviation above the next closest value. This resulted in a total of 5 cases being removed and was visualized using $\bar{X}$ plots (Table G-3, Figure G-2).

Table G-3 Anesthesia duration outliers

Procedure	Anesthesia duration (LCL)		Anesthesia duration (UCL)		Total outliers (%)	
CABG	1	(< 16)	0	(> 147)	1	(0.2%)
AVR	1	(< 22)	1	(> 121)	2	(1.3%)
MVR	1	(< 12)	0	(> 135)	1	(1.2%)
END	0	(< 15)	0	(> 76)	0	(0.0%)
HAT	0	(< 0)	0	(> 103)	0	(0.0%)
HYSD	0	(< 0)	1	(> 44)	1	(0.6%)

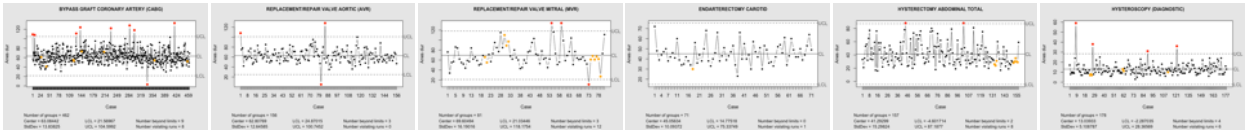


Figure G-2 $\bar{X}$ plots for anesthesia durations

Finally, outliers were identified within the other parts of surgery: preparation, wrap-up, and turnover. To remove the case as an outlier based on these values, as with the anesthesia outliers, the duration had to be both above the UCL and more than a full standard deviation above the next closest value. This resulted in a total of 27 cases being removed (Table G-4). $\bar{X}$ plots were used to visualize and identify these outliers (Figure G-3).

Table G-4 Preparation, wrap-up and turnover duration outliers

Procedure	Prep duration		Wrap-up duration		Turnover duration		Total outliers (%)	
CABG	4	(> 16)	4	(> 45)	4	(> 122)	12	(2.6%)
AVR	1	(> 15)	0	(> 51)	1	(> 154)	2	(1.3%)
MVR	0	(> 16)	1	(> 48)	1	(> 145)	2	(2.5%)
END	0	(> 12)	0	(> 54)	1	(> 49)	1	(1.4%)
HAT	0	(> 21)	1	(> 60)	1	(> 90)	2	(1.3%)
HYSD	0	(> 15)	3	(> 21)	5	(> 88)	8	(4.4%)

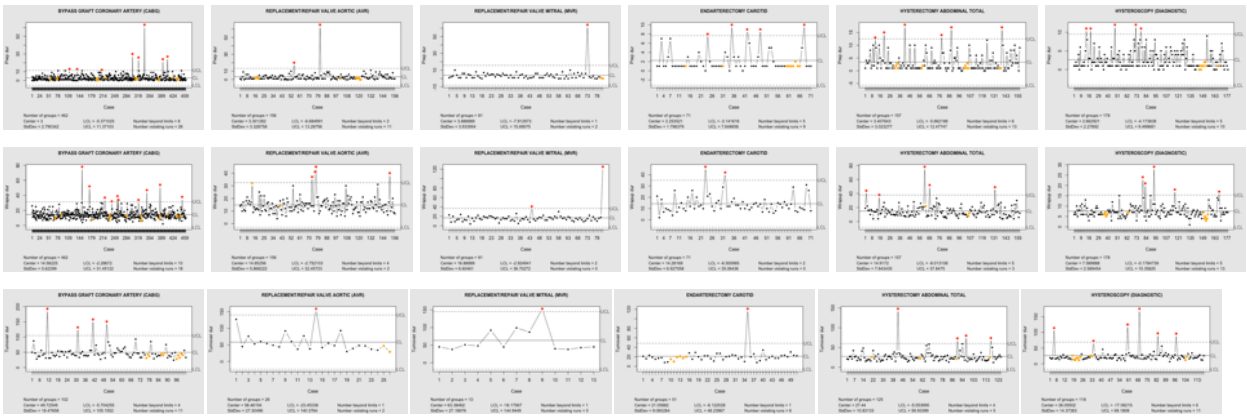


Figure G-3 $\bar{X}$ plots for preparation, wrap-up, and turnover durations

A summary of the cases that were removed due to being outliers caused by special conditions is shown in Table C-2. It sums all cases that were eliminated from each of the difference procedures. In total, 17 cases were eliminated due to errors and 47 due to outliers.

G.3 Data Exploration

Data exploration was needed prior to creating the models in order to determine how different variables interact and impact the outcome. In this instance, the variables to consider include the surgery, room, anesthesia, and patient variables (Table G-1). A breakdown of the variables is shown in Table G-5.

Table G-5 Surgical duration variables breakdown

Variable	**Factors**	**Values**					
ASA Score	5	1, 2, 3, 4, 5					
Patient Status	3	Day surgery (DS), Inpatient (INPT), Same day admittance (SDA)					
Age (min-max)	--	20-91					
Sex	2	Female, Male					
Room Description	11	01, 02, 03, 04, 07, 08, 09, 11, 12, 14, 15					
Day of the Week	5	Monday—Friday					
Case Service	3	**Cardiac**			**Vascular**	**OBGYN**	
Primary Procedure Description	6	**CABG**	**AVR**	**MVR**	**END**	**HAT**	**HYSD**
Scheduled Main Surgeon	37	8	8	8	2	23	21
Scheduled Anesthesiologist	55	21	19	16	32	37	38
Anesthesia Type	22	6	6	4	3	14	9
# of Procedures (min-max)	--	1-3	1-3	1-3	1-2	1-7	1-4
Surgeon Estimate (min-max	--	60-345	240-390	240-465	120-270	60-375	45-105
mean		299	301	301	205	170	62
median, mode)		300, 300	300, 300	300, 300	195, 195	165, 120	60, 60

Also, it may be beneficial to break the case duration into the separate parts of surgery (preparation, anesthesia, procedure, wrap-up, turnover) as the variables may affect each section differently and the durations may be more accurately predicted separately. Analysis was done to assess (1) correlation of the surgery parts to the overall case duration, (2) shape and spread of the duration distributions for each procedure, and (3) which variables to include in the model.

G.3.1 Correlation

The correlation between the case parts and total case duration was assessed using a correlation matrix for all cases together (Figure G-4) and separated by procedure (Figure G-5).

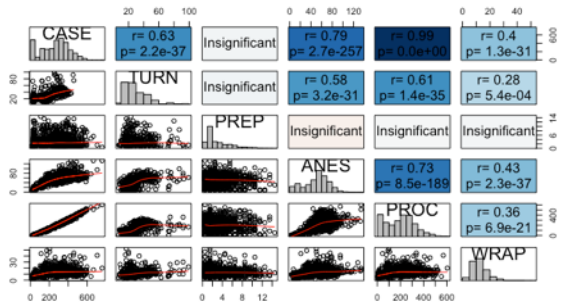


Figure G-4 Correlation matrix for all cases

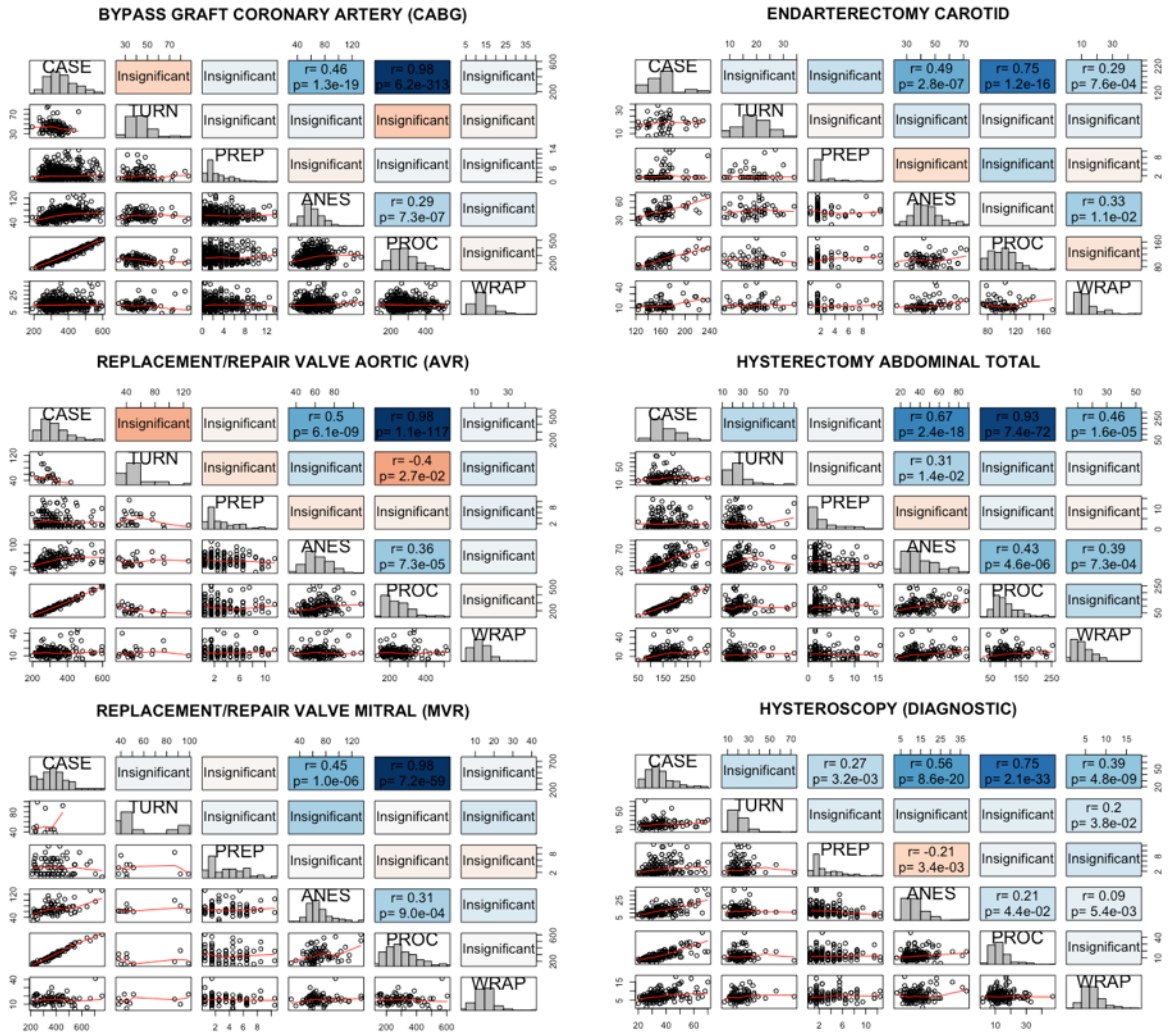


Figure G-5 Correlation matrix for each procedure

The correlation matrix includes three elements in total: a scatter plot with locally weighted smoothing (LOESS) (lower panel), histograms (diagonal panel), and Spearman correlation with p-values (upper panel). The upper panel is coloured based on significance—blue indicates a positive correlation, white for no correlation, and red for a negative correlation—and if the p-value is greater than 0.05 it shows as "insignificant". The LOESS method of fitting a smooth curve and Spearman correlation method were used to account for possibly non-parametric data.

With cases separated by procedure, there are three parts of surgery which consistently have correlation to the full case duration (and to each other): anesthesia, procedure, and wrap-up durations. Procedure duration has the highest positive correlation to case duration (0.75-0.98) with anesthesia duration being the next highest (0.45-0.67) for all procedures. With both these parts of surgery having a positive correlation to the total case duration, it is unsurprising that procedure and anesthesia also have a positive correlation to each other (0.21-0.43) for all procedures but END. The wrap-up duration is only significant for non-cardiac surgeries (END, HAT, HYSD) and is positively correlated to case duration (0.29, 0.46, 0.39) and anesthesia duration (0.33, 0.39, 0.09). All other interactions have only a single procedure for which the correlation is significant.

When all the cases are assessed together there is significant positive correlation between all surgery parts including total case duration, with the exception of preparation duration which has no significant correlation to any other section. There was no correlation for preparation when separated by procedure either, so preparation duration has no relation to any other part of surgery or dependence on procedure type. Alternatively, as turnover time did not have significant correlation for most cases when separated by procedure but did correlate when assessed with all cases, it shows that the procedure type does have a significant impact on turnover duration.

G.3.2 Duration Distributions

With correlation assessed, the next step was to assess the shape of each duration distribution. Based on the literature [34], it was expected that a lognormal distribution would best suit the duration values. This was visualized using histogram/density plots and lognormal-transformed quantile-quantile (QQ) plots—both for the case duration as a whole (Figure G-6) and for different surgical parts (Figure G-7)—and tested using a Shapiro-Wilk test (Table G-6).

Based on the QQ plots, the case duration distribution for each procedure fit well to lognormal, though outliers on the high end of the durations tend to skew from the trend. A similar spread is seen for some of the surgical parts: procedure, anesthesia, and wrap-up durations. The procedure and anesthesia durations match the curve for the lognormal distribution even more closely than the full case duration, with only mild deviations in the upper outliers. The wrap-up duration, though it mostly fits well to the lognormal distribution, shows more deviation at both ends and often a steeper slope than expected in the QQ plots such as with the CABG, END, and HYSD procedures. The preparation durations do not fit well to a normal or lognormal distribution.

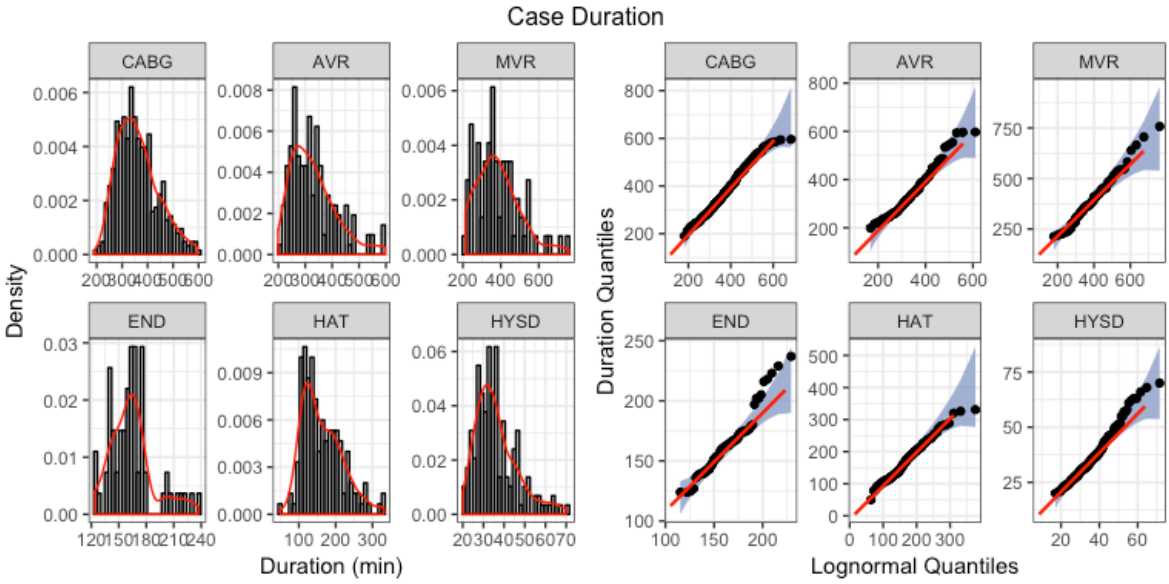


Figure G-6 Case duration distributions

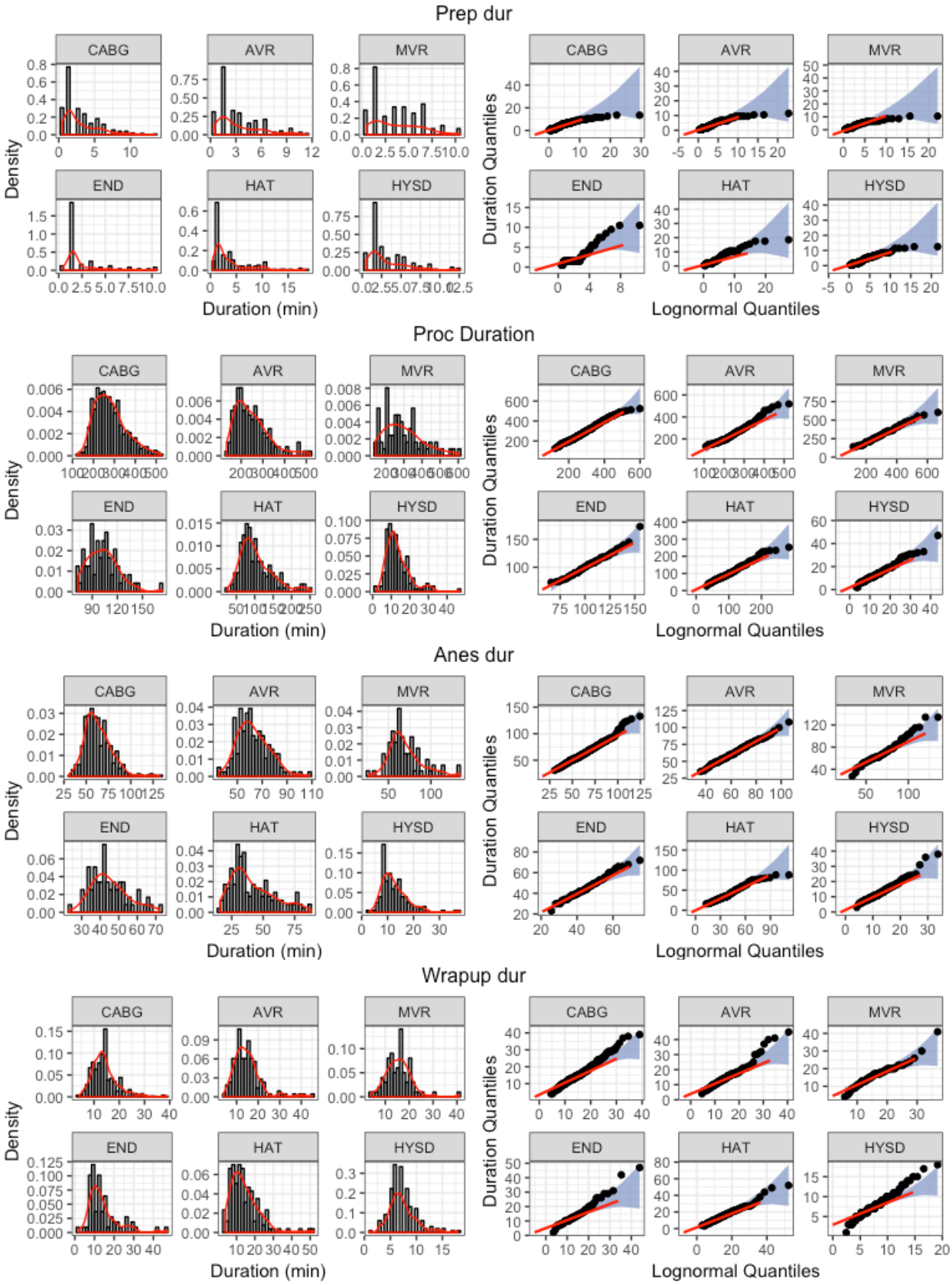


Figure G-7 Surgery parts duration distributions

The Shapiro-Wilk test was used to further justify the distribution shapes. The test was completed for each duration with the original data and after being lognormally transformed. The results are shown in Table G-6. A p-value of greater than 0.05 indicates that the assumption of a normal distribution (or lognormal, for the transformed data) could not be dismissed. With the original data, the assumption that the distribution was normal was rejected for all but one duration (END, anesthesia). After transforming the data, the lognormal distribution was still rejected for all preparation duration and all but one wrap-up duration (HAT) but was not rejected for the majority of anesthesia and procedure durations.

Table G-6 Normal vs Lognormal Shapiro-Wilk Test

Procedure	Duration	Normal p-value	Lognormal p-value
CABG	Preparation	3.65E-21	5.39E-14
	Anesthesia	6.19E-12	0.051505
	Procedure	6.41E-10	0.069601
	Wrap-up	1.01E-15	0.000476
	Case	1.4E-08	0.032119
AVR	Preparation	2.89E-11	3.53E-07
	Anesthesia	0.014565	0.91197
	Procedure	1.61E-08	0.003839
	Wrap-up	3.47E-11	0.045742
	Case	8.38E-08	0.001854
MVR	Preparation	4.75E-05	3.57E-05
	Anesthesia	0.000191	0.089918
	Procedure	0.002164	0.439039
	Wrap-up	0.000209	0.005584
	Case	0.001966	0.419931
END	Preparation	8.62E-12	4.57E-09
	Anesthesia	0.072982	0.80006
	Procedure	0.009995	0.550684
	Wrap-up	1.87E-07	0.041524
	Case	0.000315	0.017599
HAT	Preparation	5.09E-15	1.72E-08
	Anesthesia	1.88E-07	0.027585
	Procedure	5.5E-08	0.499293
	Wrap-up	1.38E-10	0.33584
	Case	8.38E-06	0.06378
HYSD	Preparation	1.72E-13	6.44E-08
	Anesthesia	1.32E-10	0.046536
	Procedure	1.38E-10	0.000835
	Wrap-up	6.92E-08	2.8E-06
	Case	6.81E-08	0.023581

p-value > 0.01

p-value > 0.05

Based on the results of the Shapiro-Wilk tests and the analysis of the visual tests, a lognormal transformation of the data is likely to create more accurate predictions for anesthesia and procedure durations in a linear model. The same is true for predicting the full case durations and wrap-up durations, but to a lesser extent as they did not fit the lognormal distribution as well as the anesthesia and procedure durations did. However, due to the wrap-up durations having a smaller range and accounting for a smaller proportion of the total case duration, a constant value specific to each procedure may also by a viable option for prediction. As the preparation durations did not fit a normal or lognormal distribution, have the smallest range, and account for the smallest proportion of the total duration, they would be best represented by constant values specific to each procedure.

G.3.3 Variables to Include in the Model

When it comes to variable selection, the type of model being generated is very important. For an explanatory model, the goal is to identify variables with a statistically significant impact on the outcome. However, for a predictive model it is not necessary for the variables to have a theoretically or statistically important impact that can be proven—the main goal is to produce accurate predictions, not to explain why or how it works. For predictive modeling, the main obstruction to including a variable lies in when the data are available. For example, including a variable about whether a surgery starts on time could aide in predicting duration but cannot be known until the surgery starts and so cannot be included in the model. Since availability and association are the main reasons for including variables in a predictive model, all of the variables discussed previously can be included. Therefore, a total of 13 variables are included, of which ten are categorical variables and three are continuous (see again Table G-5 for details on the factor levels for each variable).

When creating the models, the variables of anesthesia type, anesthesiologist, surgeon, and room number need to be given extra consideration as they have some values which occur very rarely and may have new values added that do not occur in the test data used to create the models (i.e., if a new surgeon begins working at the hospital). Therefore, we must ensure that these factors will not cause overfitting and that the model will not fail if a new value is added.

G.4 Model Generation

Three models were created to be compared against the current slating method of surgeon estimates. The three types of models created were (1) a simple moving average (SMA) model, (2) a linear mixed effects (LME) model for which two variations were completed, and (3) a random forest (RF) model. For the models which required training (LME and RF), the data were randomly separated using R [52] such that approximately 80% of the data were assigned for training and the remaining data for testing the models.

G.4.1 Simple Moving Average Model

Using the R "dplyr" package [53], the surgeries were grouped by procedure and arranged by date. The "zoo" package [54] was then used to calculate a moving average using the previous 19 cases as the predicted duration for the next case. As 19 values were required to create the moving average, the first 19 cases for each procedure had no prediction; estimates were created for all other cases.

G.4.2 Linear Mixed Effects Model

Before creating the LME model, the input variables were defined as either fixed or random effects. Fixed variables include the continuous variables of surgeon estimate, number of procedures, and age, as well as the categorical variables of sex, patient status, ASA score, and

day of the week. The random variables are case service, room number, anesthesiologist, and anesthesia type, while primary procedure and surgeon are random nested variables within case service. No interactions between variables were included and, for the random effects, only random intercepts were included, not random slopes.

Two different models were created using this method, one which used only a single LME model with the total case duration as the output (LME-case) and another which used a sum of LME models and constant values with the durations of each part of the surgery as outputs (LME-part). These two variations were created to determine if splitting the case into its parts would improve the estimate accuracy or if the less complex method of only using the case duration is adequate. Based on the data exploration completed, the preparation and wrap-up durations were calculated as constant values for each procedure while the anesthesia, procedure, and total case durations were found using LME models with log-transformed duration data.

The training data were used to create the models. To calculate the constant values, the "dplyr" package [53] was used to group surgeries by procedure and calculate the mean preparation and wrap-up durations for each procedure. The LME models for anesthesia, procedure, and case durations were created using the "lme4" package [55]. The "dplyr" package was again used to assign the appropriate constant values to the test data for both preparation and wrap-up durations. Core R code from the "stats" package [52] was used to predict the anesthesia, procedure, and case durations for the test data using the LME models and then the results were back-transformed. The predicted case durations constitute the first variation of LME model results (LME-case) to be compared against the surgeon estimate, while the sum of the predictions for preparation, anesthesia, procedure, and wrap-up durations constitute the second (LME-part).

G.4.3 Random Forest Model

The first step in creating the RF model was to group the uncommon factors into a new factor named "other" in order to prevent failure in the event of a new factor level being used that was not in the training data. The "dplyr" [53] and "forcats" packages [56] were used to form these new groups. Four of the thirteen variables required the new level to be added: room number (11 levels, 3 with <20 occurrences), surgeons (37 levels, 8 with <3 occurrences), anesthesiologist (51 levels, 23 with <12 occurrences), and anesthesia type (22 levels, 4 with <2 occurrences).

The RF model was created with the training data using the "randomForest" package [57], with case duration as the output and all 13 variables included. The "randomForest" package was also used to determine the optimal number of variables randomly sampled for splitting at each tree node which was found to be four (m_{try}, Figure G-8), as well as to measure and graph variable importance (Figure G-9). As with the LME models, core R code from the "stats" package [52] was then used to predict the case durations for the test data using the RF model and then the results were back-transformed.

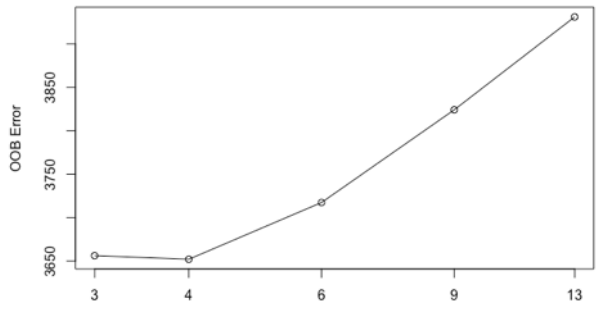


Figure G-8 Optimal value of m_{try} for randomForest

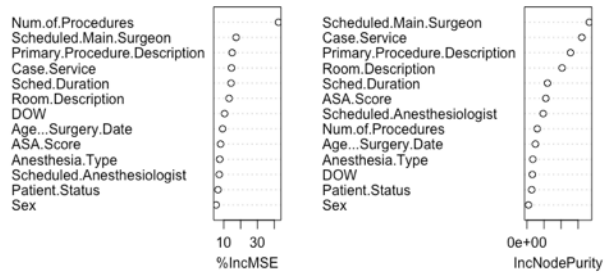


Figure G-9 Variable importance dot-chart

G.5 Model Testing

Two rounds of testing were completed for the models. The first was done using the 20% of remaining cases which had been separated from the model training data, after filtering for only cases which also had a SMA prediction (eliminating any predictions which were provided for the first 19 cases of each procedure). The second round of testing used new data for cases occurring from Apr'18 to Jul'19 to test how well the models perform on new, unfiltered data over an extended period of time. The new data were edited for errors but not for outliers. The number of cases compared for each procedure in the two rounds of testing is shown in Table G-7.

Table G-7 Cases included in each round of testing

Values	Total	CABG	AVR	MVR	END	HAT	HYSD
Round 1	210	88	31	7	12	28	44
Round 2	1314	498	209	104	82	227	194
Total	1524	586	240	111	94	255	238

In each round of testing the prediction errors were compiled and compared against the surgeon errors. To account for the surgeon estimates including turnover times while the new models did not, the average turnover time for each procedure from the associated data set was subtracted from the surgeon estimate for cases which were not the only or last case of the day.

Each of the models were compared against the surgeon estimate for all cases combined and for each procedure individually. The values used to compare the estimation methods are percent of cases which finished on-time, percent of cases improved from the surgeon estimate, process sigma value, error distribution values (mean, standard deviation, minimum, maximum), and error sums (overtime, undertime, sum of absolute error, sum of error). Values which showed an improvement from the surgeon estimate are filled green and the best result for each value is coloured more darkly. Scatterplots of the estimates vs. actual durations were also assessed for all cases combined and for each procedure.

G.5.1 Model Testing – Round 1, original data

The values used to compare the models based on the predictions generated in the first round of testing are shown in Table G-8 for all cases combined and Table G-9 to Table G-14 for each procedure (CABG, AVR, MVR, END, HAT, HYSD). The predictions are visualized against the actual durations using scatter plots for all cases/models combined (Figure G-10) and separated by model and procedure (Figure G-11). Due to some procedures having few cases (MVR, END) the majority of analysis completed in Section 3.3.4.2 is not based on this round of testing.

Table G-8 Round 1, All Cases: Comparison of surgery duration estimate methods

Values	Surgeon	SMA	LME-case	LME-part	RF
Total cases	210	210	210	210	210
Improved cases (%)	--	63.3	68.1	66.7	66.7
On-time cases, yield (%)	25.2	31.9	36.7	33.8	32.9
Short term process sigma (1.5 shift)	0.83	1.03	1.16	1.08	1.06
Mean error	-38.4	-0.9	-4.7	-5.5	0.0
Standard deviation error	76.3	68.0	55.1	55.1	53.8
Minimum error	-343	-265	-219	-226	-219
Maximum error	94	160	159	159	163
Sum of undertime (+ error)	2136	4931	3504	3439	3924
Sum of overtime (- error)	-10203	-5112	-4486	-4587	-3915
Sum of |error|	12339	10043	7990	8026	7839
Sum of error	-8066	-180	-982	-1148	9

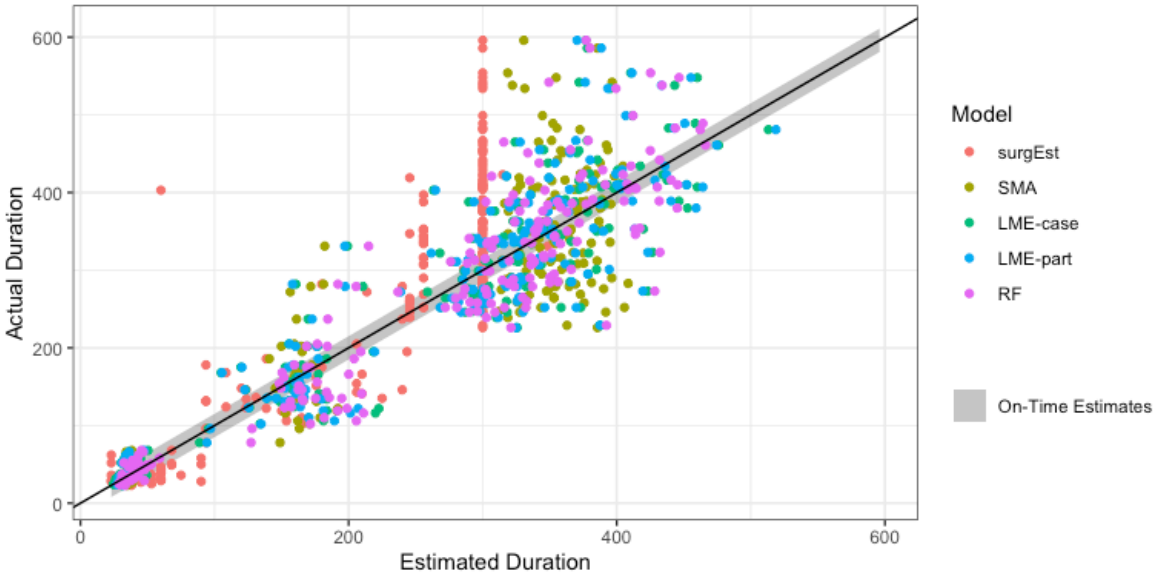


Figure G-10 Round 1, All Cases: Estimated duration vs actual duration scatter plot

Table G-9 Round 1, CABG: Comparison of surgery duration estimate methods

Values	Surgeon	SMA	LME-case	LME-part	RF
Total cases	88	88	88	88	88
Improved cases (%)	--	63.6	70.5	70.5	70.5
On-time cases, yield (%)	13.6	17.0	19.3	21.6	21.6
Short term process sigma (1.5 shift)	0.40	0.55	0.63	0.71	0.71
Mean error	-69.0	6.3	1.2	0.5	0.3
Standard deviation error	76.8	75.5	62.2	62.0	59.0
Minimum error	-343	-216	-208	-198	-206
Maximum error	71	154	159	159	163
Sum of undertime (+ error)	457	2876	2139	2097	1952
Sum of overtime (- error)	-6532	-2323	-2034	-2057	-1921
Sum of |error|	6989	5199	4174	4154	3873
Sum of error	-6076	552	105	41	31

Table G-10 Round 1, AVR: Comparison of surgery duration estimate methods

Values	Surgeon	SMA	LME-case	LME-part	RF
Total cases	31	31	31	31	31
Improved cases (%)	--	61.3	61.3	61.3	54.8
On-time cases, yield (%)	29.0	22.6	22.6	16.1	12.9
Short term process sigma (1.5 shift)	0.95	0.75	0.75	0.51	0.37
Mean error	-57.9	-13.8	-21.4	-22.5	-9.8
Standard deviation error	87.5	92.8	65.2	66.6	65.4
Minimum error	-296	-265	-219	-226	-219
Maximum error	53	91	71	73	73
Sum of undertime (+ error)	186	836	422	433	584
Sum of overtime (- error)	-1981	-1265	-1086	-1130	-889
Sum of |error|	2167	2101	1508	1563	1474
Sum of error	-1796	-429	-664	-697	-305

Table G-11 Round 1, MVR: Comparison of surgery duration estimate methods

Values	Surgeon	SMA	LME-case	LME-part	RF
Total cases	7	7	7	7	7
Improved cases (%)	--	85.7	85.7	85.7	85.7
On-time cases, yield (%)	0.0	14.3	14.3	14.3	14.3
Short term process sigma (1.5 shift)	-4.50	0.43	0.43	0.43	0.43
Mean error	-117.7	-35.7	-34.8	-31.9	-43.2
Standard deviation error	100.3	105.8	91.1	89.4	94.5
Minimum error	-242	-154	-170	-165	-193
Maximum error	74	160	100	99	95
Sum of undertime (+ error)	74	170	148	143	130
Sum of overtime (- error)	-898	-420	-391	-366	-433
Sum of |error|	972	590	539	509	563
Sum of error	-824	-250	-243	-223	-302

Table G-12 Round 1, END: Comparison of surgery duration estimate methods

Values	Surgeon	SMA	LME-case	LME-part	RF
Total cases	12	12	12	12	12
Improved cases (%)	--	50.0	58.3	58.3	50.0
On-time cases, yield (%)	50.0	50.0	50.0	50.0	33.3
Short term process sigma (1.5 shift)	1.50	1.50	1.50	1.50	1.07
Mean error	22.5	-7.2	-7.7	-7.1	3.5
Standard deviation error	34.2	30.8	29.5	30.0	31.4
Minimum error	-26	-76	-67	-64	-53
Maximum error	94	27	30	31	50
Sum of undertime (+ error)	302	91	82	89	176
Sum of overtime (- error)	-31	-178	-174	-175	-134
Sum of |error|	333	269	256	264	309
Sum of error	270	-86	-92	-86	42

Table G-13 Round 1, HAT: Comparison of surgery duration estimate methods

Values	Surgeon	SMA	LME-case	LME-part	RF
Total cases	28	28	28	28	28
Improved cases (%)	--	42.9	50.0	46.4	46.4
On-time cases, yield (%)	14.3	10.7	32.1	21.4	10.7
Short term process sigma (1.5 shift)	0.43	0.26	1.04	0.71	0.26
Mean error	-1.3	6.4	5.2	2.0	20.1
Standard deviation error	48.5	62.7	56.7	56.3	53.4
Minimum error	-102	-149	-132	-134	-116
Maximum error	90	72	104	85	100
Sum of undertime (+ error)	542	810	662	623	939
Sum of overtime (- error)	-579	-629	-517	-567	-375
Sum of |error|	1121	1439	1179	1190	1314
Sum of error	-37	181	145	56	564

Table G-14 Round 1, HYSD: Comparison of surgery duration estimate methods

Values	Surgeon	SMA	LME-case	LME-part	RF
Total cases	44	44	44	44	44
Improved cases (%)	--	77.3	79.5	75.0	81.8
On-time cases, yield (%)	50.0	79.5	84.1	79.5	86.4
Short term process sigma (1.5 shift)	1.50	2.33	2.50	2.33	2.60
Mean error	9.0	-3.4	-5.3	-5.4	-0.5
Standard deviation error	19.6	12.1	9.4	9.7	9.1
Minimum error	-39	-32	-30	-30	-23
Maximum error	62	12	13	9	19
Sum of undertime (+ error)	577	149	51	54	143
Sum of overtime (- error)	-181	-296	-284	-292	-163
Sum of |error|	758	445	335	345	306
Sum of error	396	-148	-233	-238	-20

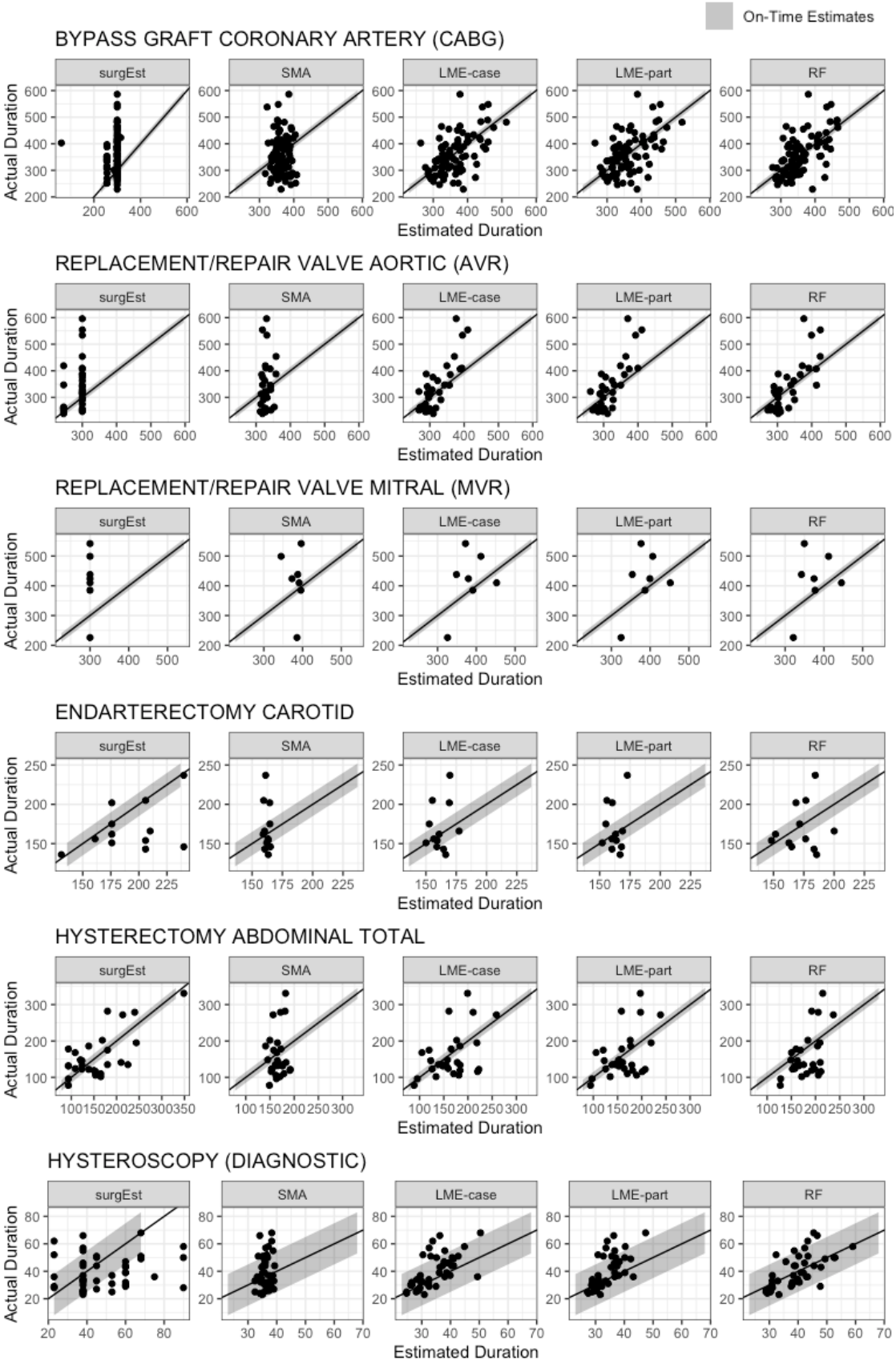


Figure G-11 Round 1, Model/Procedure: Estimated duration vs actual duration scatter plots

G.5.2 Model Testing – Round 2, new data

The values used to compare the models based on the predictions generated in the second round of testing are shown in Table G-15 for all cases combined and Table G-16 to Table G-21 for each of the procedures (CABG, AVR, MVR, END, HAT, HYSD). The predictions are visualized against the actual durations using scatter plots for all cases/models combined (Figure G-12) and separated by model and procedure (Figure G-13). A summary and analysis of the results is completed in Section 3.3.4.2.

Table G-15 Round 2, All Cases: Comparison of surgery duration estimate methods

Values	Surgeon	SMA	LME-case	LME-part	RF
Total cases	1314	1314	1314	1314	1314
Improved cases (%)	--	63.4	74.7	74.3	70.2
On-time cases, yield (%)	19.0	27.5	31.0	30.6	30.5
Short term process sigma (1.5 shift)	0.62	0.90	1.00	0.99	0.99
Mean error	-65.0	-0.9	-21.5	-22.5	-23.1
Standard deviation error	95.1	83.7	70.6	70.3	72.8
Minimum error	-528	-444	-409	-418	-439
Maximum error	206	276	253	245	239
Sum of undertime (+ error)	10245	38293	18079	17269	18048
Sum of overtime (- error)	-95643	-39447	-46280	-46844	-48444
Sum of |error|	105888	77740	64360	64113	66492
Sum of error	-85398	-1154	-28201	-29574	-30396

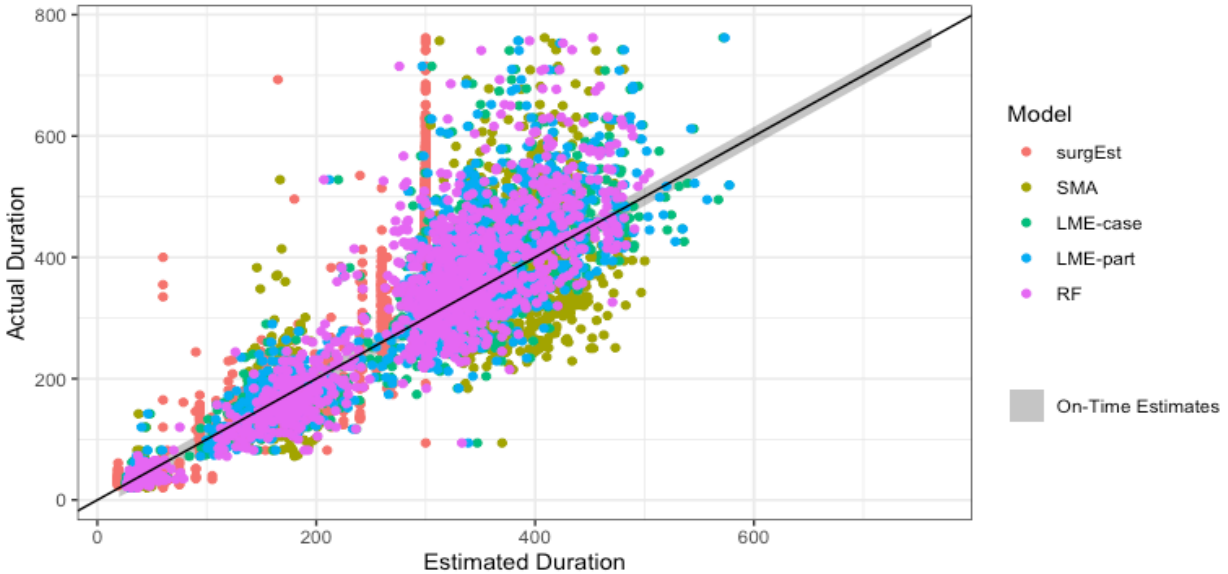


Figure G-12 Round 2, All Cases: Estimated duration vs actual duration scatter plot

Table G-16 Round 2, CABG: Comparison of surgery duration estimate methods

Values	Surgeon	SMA	LME-case	LME-part	RF
Total cases	498	498	498	498	498
Improved cases (%)	--	73.1	85.7	85.9	79.3
On-time cases, yield (%)	5.0	16.3	17.9	17.7	18.3
Short term process sigma (1.5 shift)	-0.14	0.52	0.58	0.57	0.59
Mean error	-110.3	-0.2	-39.5	-40.8	-47.0
Standard deviation error	88.3	89.3	76.0	76.0	77.5
Minimum error	-453	-337	-343	-349	-363
Maximum error	206	276	253	245	239
Sum of undertime (+ error)	993	17266	5703	5489	4866
Sum of overtime (- error)	-55918	-17355	-25389	-25791	-28255
Sum of |error|	56911	34621	31092	31281	33121
Sum of error	-54925	-89	-19686	-20302	-23389

Table G-17 Round 2, AVR: Comparison of surgery duration estimate methods

Values	Surgeon	SMA	LME-case	LME-part	RF
Total cases	209	209	209	209	209
Improved cases (%)	--	60.8	73.7	72.7	68.4
On-time cases, yield (%)	10.0	11.5	14.4	15.8	15.8
Short term process sigma (1.5 shift)	0.22	0.30	0.44	0.50	0.50
Mean error	-88.1	-5.7	-41.3	-41.6	-34.9
Standard deviation error	110.3	111.9	88.3	87.6	86.2
Minimum error	-528	-444	-372	-373	-390
Maximum error	108	222	145	146	117
Sum of undertime (+ error)	1675	8520	2924	2838	3252
Sum of overtime (- error)	-20089	-9717	-11559	-11525	-10551
Sum of |error|	21763	18236	14483	14363	13802
Sum of error	-18414	-1197	-8635	-8687	-7299

Table G-18 Round 2, MVR: Comparison of surgery duration estimate methods

Values	Surgeon	SMA	LME-case	LME-part	RF
Total cases	104	104	104	104	104
Improved cases (%)	--	62.5	65.4	64.4	69.2
On-time cases, yield (%)	7.7	11.5	6.7	4.8	13.5
Short term process sigma (1.5 shift)	0.07	0.30	0.00	-0.16	0.40
Mean error	-104.5	-0.1	0.0	-3.2	-26.7
Standard deviation error	116.3	121.4	92.0	91.9	101.2
Minimum error	-462	-353	-409	-418	-439
Maximum error	82	205	162	158	162
Sum of undertime (+ error)	763	5004	3747	3586	2628
Sum of overtime (- error)	-11633	-5016	-3750	-3918	-5407
Sum of |error|	12397	10020	7497	7505	8035
Sum of error	-10870	-13	-4	-332	-2779

Table G-19 Round 2, END: Comparison of surgery duration estimate methods

Values	Surgeon	SMA	LME-case	LME-part	RF
Total cases	82	82	82	82	82
Improved cases (%)	--	69.5	70.7	72.0	72.0
On-time cases, yield (%)	24.4	39.0	41.5	41.5	41.5
Short term process sigma (1.5 shift)	0.81	1.22	1.28	1.28	1.28
Mean error	13.5	2.2	-6.9	-6.6	1.3
Standard deviation error	50.8	26.7	28.0	27.6	26.5
Minimum error	-109	-62	-80	-84	-72
Maximum error	107	45	52	49	48
Sum of undertime (+ error)	2259	1010	639	627	934
Sum of overtime (- error)	-1151	-827	-1208	-1170	-826
Sum of |error|	3410	1838	1848	1797	1761
Sum of error	1108	183	-569	-543	108

Table G-20 Round 2, HAT: Comparison of surgery duration estimate methods

Values	Surgeon	SMA	LME-case	LME-part	RF
Total cases	227	227	227	227	227
Improved cases (%)	--	44.5	57.3	54.6	53.3
On-time cases, yield (%)	26.9	22.0	30.4	29.5	30.8
Short term process sigma (1.5 shift)	0.88	0.73	0.99	0.96	1.00
Mean error	-15.3	-0.2	4.8	4.0	9.0
Standard deviation error	44.7	67.0	49.1	48.2	48.9
Minimum error	-169	-361	-308	-316	-321
Maximum error	128	109	151	139	119
Sum of undertime (+ error)	2402	5512	4524	4310	4980
Sum of overtime (- error)	-5880	-5567	-3431	-3411	-2946
Sum of |error|	8281	11079	7955	7721	7926
Sum of error	-3478	-55	1093	899	2033

Table G-21 Round 2, HYSD: Comparison of surgery duration estimate methods

Values	Surgeon	SMA	LME-case	LME-part	RF
Total cases	194	194	194	194	194
Improved cases (%)	--	61.3	74.7	75.3	68.6
On-time cases, yield (%)	59.3	83.5	91.8	90.2	82.0
Short term process sigma (1.5 shift)	1.73	2.47	2.89	2.79	2.41
Mean error	6.1	0.1	-2.1	-3.1	4.8
Standard deviation error	20.1	15.0	12.6	12.5	12.2
Minimum error	-60	-104	-94	-96	-45
Maximum error	71	28	27	12	60
Sum of undertime (+ error)	2154	982	543	419	1388
Sum of overtime (- error)	-973	-964	-942	-1028	-459
Sum of |error|	3127	1945	1485	1447	1847
Sum of error	1181	18	-399	-609	930

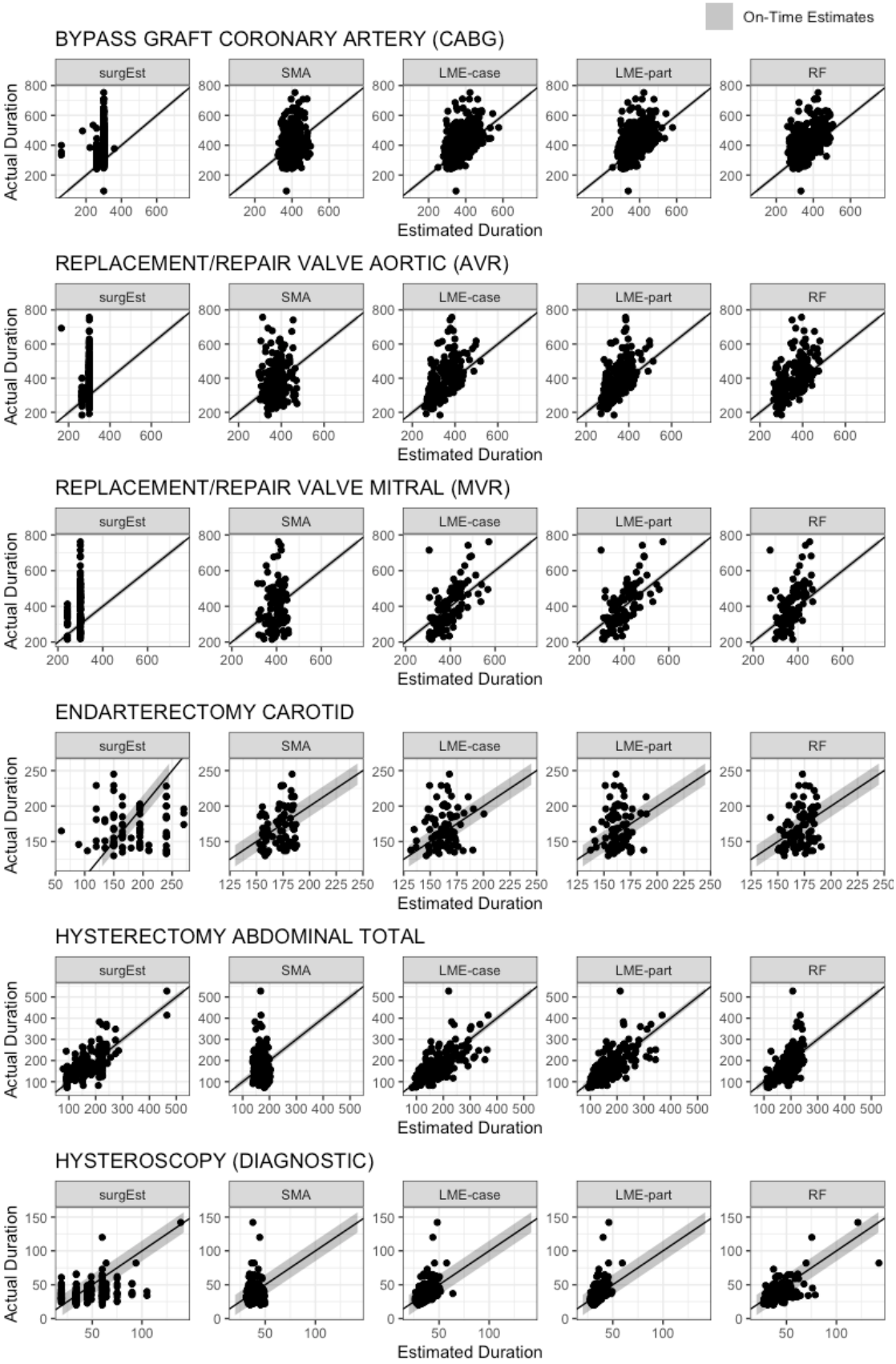


Figure G-13 Round 2, Model/Procedure: Estimated duration vs actual duration scatter plots

G.6 Model Code/Functions

To generate predictions, three steps were taken. First, functions were created which include all data needed to generate predictions. Second, data were manipulated as needed. And finally, the data and functions were used to generate predictions that were combined to one data-frame, which could then be used to compare the results from each method against the surgeon estimates.

G.6.1 Functions

Functions were created so that they could be used multiple times and generate predictions for different data, dependent on the data read into it. Within each function is the information needed to create the model and the code to generate predictions; the predictions are returned. The models and associated functions used to create them are: SMA (*SMAmodel* function), LME-case (*LMEmodel* function), LME-part (2 *LMEmodel* functions + 2 *constantModel* functions), and RF (*RFmodel* function). Figure G-14 to Figure G-17 show the coding used to create these functions.

```
SMAmodel <- function(data) {                                    #function to create SMA predictions
  SMApredict <- data_frame()
  procnames <- c("BYPASS GRAFT CORONARY ARTERY (CABG)",         #list of procedure names
                 "REPLACEMENT/REPAIR VALVE AORTIC (AVR)",
                 "REPLACEMENT/REPAIR VALVE MITRAL (MVR)",
                 "ENDARTERECTOMY CAROTID",
                 "HYSTERECTOMY ABDOMINAL TOTAL",
                 "HYSTEROSCOPY (DIAGNOSTIC)")
  n = length(procnames)                                         #number of procedures (6)
  for (i in 1:n) {                                              #for loop, 1 to length of "procnames" (6)
    #filter and order data
    simpdata <- data %>%
      filter(`Primary.Procedure.Description`==procnames[i]) %>%   #filter for primary procedure in "procnames"
      arrange(`Case.Date`)                                        #arrange by date

    #calculate simple moving average predictions
    SMAmodel <- simpdata %>%                                  #make new column (SMA) with rolling average
      mutate(SMA = lag(rollapply(simpdata$`Case.Duration`,19,mean,align="right",fill=NA)))

    #add predictions to dataframe
    SMApredict <- rbind(SMApredict, SMAmodel)          #bind predictions for each procedure to one dataframe
  }
  return(SMApredict)
}

#SMAmodel(data)
  #model to create moving average estimates; only works for 6 procedures* (CABG, AVR, MVR, END, HAT, HYSD)
  #    *change values in "procnames" to filter and get values for different procedures
#data:
  #past procedure data, must include columns `Primary.Procedure.Description`, `Case.Date`, and `Case.Duration`
```

Figure G-14 SMAmodel Function

```
LMEmodel <- function(traindata, testdata, trainDur, predName) {   #function to create LME models
  LMEtraindata <- traindata                #assign lines with (1) as training data = ~80% of all data
  LMEtestdata <- testdata                  #assign lines with (2) as testing data = ~20% of all data

  LMEmodel <- lmer(log(trainDur) ~         #linear mixed effects model for log-transformed duration data

                 #FIXED
                 Sched.Duration +                  #num
                 Age...Surgery.Date +              #num
                 Num.of.Procedures +               #num
                 Sex +                             #factor w/ 2
                 Patient.Status +                  #factor w/ 3
                 ASA.Score +                       #factor w/ 5+NA
                 DOW +                             #factor w/ 5

                 #RANDOM
                 (1|Case.Service) +                     #factor
                 (1|Primary.Procedure.Description) +    #factor
                 (1|Room.Description) +                 #factor
                 (1|Scheduled.Anesthesiologist) +       #factor
                 (1|Anesthesia.Type) +                  #factor

                 #RANDOM - NESTED within case.service
                 (1|Case.Service:Scheduled.Main.Surgeon) ,   #factor

               data = LMEtraindata)                            #data used to train the model

  LMEpredict <- exp(predict(LMEmodel, LMEtestdata, allow.new.levels = TRUE))  #get predictions for testdata
  LMEpredict <- data.frame(predict = LMEpredict)               #put predictions into a dataframe
  colnames(LMEpredict) <- c(predName)
  return(LMEpredict)   #returns dataframe with predictions for case part
}

#LMERmodel(traindata, testdata, trainDur, predName)
  #linear mixed effects model with fixed, random, and nested variables
#traindata:
  #data used to train the model; must include all variables in the model
#testdata:
  #data used to test the model (predictions generated for); must include all variables in the model
#trainDur:
  #the column from the training data which matches the duration variable which the model is to predict
#predName:
  #name for the prediction of the case part
```

Figure G-15 LMEmodel Function

```
constantModel <- function(traindata, testdata, casePartDur, predName) {   #function to find constant values
  CONSTtraindata <- traindata                #assign lines with (1) as training data = ~80% of all data
  CONSTtestdata <- testdata                  #assign lines with (2) as testing data = ~20% of all data

  constants <- CONSTtraindata %>%                       #code used to find constant values (mean durations)
    group_by(Primary.Procedure.Description) %>%         #group by procedure
    summarise(prediction = mean(!! sym(casePartDur)))  #calculate mean duration of case part for each procedure
  colnames(constants)[2] <- c(predName)

  CONSTpredict <- left_join(CONSTtestdata,constants,   #add constant values as predicted duration for case part
                            by = "Primary.Procedure.Description")                              #based on procedure
  CONSTpredict <- CONSTpredict %>% select(predName)
  return(CONSTpredict)                                  #returns dataframe with predictions for case part
}

#constantModel(traindata, testdata, casePartDur, predName)
  #calculated mean duration for case parts and uses this constant value as the prediction
#traindata:
  #data used to train the model; must include all variables in the model
#testdata:
  #data used to test the model (predictions generated for); must include all variables in the model
#casePartDur:
  #column title/duration variable which the constant is being calculated for
#predName:
  #name for the prediction of the case part
```

Figure G-16 ConstantModel Function

```
RFmodel <- function(traindata, testdata, testcasenum){
  #determine optimum mtry (number of predictors to be randomly sampled at each split)
  bestmtry <- tuneRF(traindata[,-1],traindata[,1],mtryStart = 13,     #determine optimum mtry
                     stepFactor = 1.5,improve = 1e-10,ntreeTry = 500)

  RFmodel <- randomForest(Case.Duration~.,data=traindata,      #random forest model (with 13 input variables)
                          mtry=min(bestmtry[2]),importance=TRUE,
                          na.action = na.exclude)

  RF <- predict(RFmodel,testdata)
  RFpredict <- data.frame(RF)
  RFpredict <- cbind(testcasenum, RFpredict)

  return(RFpredict)
}

#RFmodel(data)
  #random forest model; only works if factor levels of train and test data match
#traindata:
  #data used to train the model; must include all variables in the model
#testdata:
  #data used to test the model (predictions generated for); must include all variables in the model
```

Figure G-17 RFmodel Function

G.6.2 Round 1: Data and Predictions

For round 1 of testing the models, the data from April 2017 to April 2018 which had been filtered and edited for errors were used. It was separated into two sets of data: training and testing (Figure G-18).

```
modeldata <- modeldata                          #filtered/cleaned data from Apr17-Apr18
set.seed(10)                                    #set so random sample will remain consistent for train/test data
ind <- sample(2,nrow(modeldata),replace = TRUE, #randomly assign each line as (1) or (2)
              prob = c(0.8,0.2))                    #(1) = 80% chance, (2) = 20% chance
traindata <- modeldata[ind == 1,]               #assign lines with (1) as training data = ~80% of all data
testdata <- modeldata[ind == 2,]                #assign lines with (2) as testing data = ~20% of all data
```

Figure G-18 Round 1: Data

The data required additional manipulation before being used in the RF model (Figure G-19). This was due to some categorical variables (*variableDesc*) having too many or very rare factor levels. After determining the maximum number of factors that should be keep, a minimum number of occurrences was determined for each variable (*variableMin*), such that any factors with fewer occurrences would be grouped together in a new category called “Other”. After this, the data were separated into training and testing data in the same manner as the original data.

```
#filter data for only required variables before using in RF model
RFdata <-  modeldata %>% select(Case.Duration,                  #select output variable - case duration
                                Room.Description,               #select all input variables for model (13 total)
                                DOW,
                                Case.Service,
                                Primary.Procedure.Description,
                                Sched.Duration,
                                Age...Surgery.Date,
                                Sex,
                                Scheduled.Main.Surgeon,
                                Scheduled.Anesthesiologist,
                                Patient.Status,
                                Anesthesia.Type,
                                ASA.Score,
                                Num.of.Procedures)

#edit data before using in RF model (too many factors/uncommon factors can create problems)
variableDesc <- RFdata %>% select(Room.Description,     #select variables with too many/uncommon factors
                                  Scheduled.Main.Surgeon,
                                  Scheduled.Anesthesiologist,
                                  Anesthesia.Type)
a <- ncol(variableDesc)                                 #count number of columns selected
variableDesc <- colnames(variableDesc)                  #get names of columns selected
variableMin <- c(20, #room                              #set minimum number of occurrences for each variable
                 3,  #surgeon                             (factors with less than min will be grouped)
                 12, #anesthesiologist
                 2)  #anes type

for (i in 1:(a)) {                                      #for each variable with uncommon factors
  RFdata <- RFdata %>% mutate(!!variableDesc[i] :=      #group uncommon factors together as new factor "other"
                                fct_lump_min(RFdata[,variableDesc[i]],     #for selected variables
                                             variableMin[i]))              #by the min occurrences set
}

#set training and testing data (same breakup of cases as used for LME models)
RFtraindata <- RFdata[ind == 1,]              #assign lines with (1) as training data = ~80% of all data
RFtestdata <- RFdata[ind == 2,]               #assign lines with (2) as testing data = ~20% of all data
```

Figure G-19 Round 1: RF Data, additional editing

Predictions were generated for each of the four models using the functions and data described. The results from each model—along with variables from the original data (date, procedure, actual duration, and surgeon estimate)—were then combined into a single data-frame (Figure G-20).

```
#use SMAmodel function to generate predictions
predictionsSMA <- SMAmodel(data = modeldata) %>% select(Case.,SMA)

#use LMEmodel function to generate case predictions
predictionsLMEcase <- cbind(testdata["Case."],
                            LMEmodel(traindata = traindata,
                                     testdata = testdata,
                                     trainDur = traindata$Case.Duration,
                                     predName = "CASEpred")) %>%
  transmute(Case. = Case.,
            LMEcase = CASEpred)

#use LMEmodel and constantModel functions to generate predictions for case parts and combine for case predictions
predictionsLMEpart <- cbind(testdata["Case."],                                          #Case#
              constantModel(traindata, testdata, "Prep.dur", "PREPpred"),         #PREP = constant
              LMEmodel(traindata, testdata, traindata$Anes.dur, "ANESpred"),      #ANES = LME
              LMEmodel(traindata, testdata, traindata$Proc.Duration, "PROCpred"), #PROC = LME
              constantModel(traindata, testdata, "Wrapup.dur", "WRAPpred")) %>%   #WRAP = constant
  transmute(Case. = Case.,
            LMEpart = PREPpred + ANESpred + PROCpred + WRAPpred)

#use RFmodel function to generate predictions
predictionsRF <- RFmodel(traindata = RFtraindata,
                         testdata = RFtestdata,
                         testcasenum = testdata[c("Case.")])


#COMBINE PREDICTIONS
predictions <- list(modeldata[c("Case.","Case.Date","Primary.Procedure.Description","Case.Duration","surgEst")],
                    predictionsSMA,
                    predictionsLMEcase,
                    predictionsLMEpart,
                    predictionsRF) %>% reduce(left_join, by = "Case.")
```

Figure G-20 Round 1: Predictions

G.6.3 Round 2: Data and Predictions

Data for the second round of model testing were from April 2018 to July 2019. These data had already been filtered in the same way as the data for the first round of testing. In order to properly use the data for the models, however, some additional work had to be done (Figure G-21). There were three variations on the new data produced: the new data alone (*newcasedata*), round one and round two data joined (*alldata*), and the new data and round one test data joined

(*alltestdata*). By joining the data with the round one data first, all original factor levels were maintained and any new factor levels from the round two data were added on after them.

```
newcasedata <- read_excel("new case values - filtered.xlsx")          #new data: APR18-JUL19 data
newcasedata <- as.data.frame(unclass(newcasedata), stringsAsFactors = TRUE)  #set columns as factors/num
newcols <- colnames(newcasedata)                                      #names of columns in new data

modeldata <- modeldata                                                #filtered/cleaned data from Apr17-Apr18
alldata <- rbind(subset(modeldata, select = newcols),newcasedata)     #APR17-JUL19 data, only columns in newdata

set.seed(10)                                    #set so random sample will remain consistent for train/test data
ind <- sample(2,nrow(modeldata),replace = TRUE, #randomly assign each line as (1) or (2)
              prob = c(0.8,0.2))                    #(1) = 80% chance, (2) = 20% chance
traindata <- modeldata[ind == 1,]               #assign lines with (1) as training data = ~80% of all data
testdata <- modeldata[ind == 2,]                #assign lines with (2) as testing data = ~20% of all data
alltestdata <- rbind(subset(testdata, select = newcols),newcasedata)  #all (new+old) data minus training data
```

Figure G-21 Round 2: Data

As in the first round of testing, the data required additional manipulation before they could be used for the RF model (Figure G-22). Unlike the LME model which allows for new factor levels, the RF model requires all factor levels for the data-frame returning predictions to be exactly the same as those available in the training data. Therefore, all factor-based variables in the data were edited such that the original factors were available (even if not in the new data) and any factors which were not present in the round one data were renamed as "Other". Once the factors were cleaned up, the new data were joined to the old to ensure that the factor levels were maintained.

```
newRFdata <- newcasedata %>% select(Case.,
                                    Case.Duration,              #select output variable - case duration
                                    Room.Description,           #select all input variables for model (13 total)
                                    DOW,
                                    Case.Service,
                                    Primary.Procedure.Description,
                                    Sched.Duration,
                                    Age...Surgery.Date,
                                    Sex,
                                    Scheduled.Main.Surgeon,
                                    Scheduled.Anesthesiologist,
                                    Patient.Status,
                                    Anesthesia.Type,
                                    ASA.Score,
                                    Num.of.Procedures)

#sort new factors = "Other" (RFmodel doesn't accept new levels; all old levels must be included even if not used)
RFdatafactors <- cbind(modeldata[c("Case.")], RFdata)           #original RFdata with case numbers
datafactors <- RFdatafactors[,sapply(RFdatafactors, is.factor)] #filter for columns that have factors
z <- ncol(datafactors)                                          #number of columns with factors
for (i in 2:z){                                                 #for each column with factors (skipping Case#)
  factordata <- newRFdata %>%                                   #filter for factors that match (new = old)
    filter(.data[[colnames(datafactors[i])]] %in% levels(datafactors[,i]))
  factorOther <- newRFdata %>%                                  #filter for factors that don't match (new!=old)
    filter(!.data[[colnames(datafactors[i])]] %in% levels(datafactors[,i])) %>%
    mutate(!!colnames(datafactors)[i] := "Other")               #change different factors to "Other"
  newRFdata <- rbind(factordata,factorOther)                    #bind matching/not matching rows
  newRFdata <- droplevels(newRFdata)                            #drop levels with no values
}
newRFdata <- rbind(RFdatafactors,newRFdata)                     #bind to original data to maintain factor levels
```

Figure G-22 Round 2: RF Data, additional editing

Predictions were generated for the second round of testing in a similar way to the first (Figure G-23). The training data remained the same, but the testing data varied. For the SMA predictions, all of the combined data were used so that the same first 19 cases would have no prediction, but all other cases would—including every value in the new data set. For the LME models, the new and original test data combined were used as the new test data. This was done to ensure that the factor levels were maintained with the new factors being added to the end, as already discussed. The data which had been edited and added to the original RF data were used to generate predictions for the RF model. These data included all cases from both data sets (the same as the data used for the SMA model, but with different factor levels). Once all predictions were generated and combined, the data-frame was filtered so that only cases from the new data (Apr'18-Jul'19) remained.

```
#use SMAmodel function to generate predictions
predictionsSMA2 <- SMAmodel(data = alldata) %>% select(Case.,SMA) #use alldata so first 19 come from old data
                                                                  #     are only cases with no prediction
#use LMEmodel function to generate case predictions
predictionsLMEcase2 <- cbind(alltestdata["Case."],                #use alltestdata so factors are same + new
                             LMEmodel(traindata = traindata,      #train with original training data (=same mode
l)
                                      testdata = alltestdata,     #train with alltestdata, new factors add to end
                                      trainDur = traindata$Case.Duration,
                                      predName = "CASEpred")) %>%
  transmute(Case. = Case.,
            LMEcase = CASEpred)

#use LMEmodel and constantModel functions to generate predictions for case parts and combine for case predictions
predictionsLMEpart2 <- cbind(alltestdata["Case."],                                        #Case#
              constantModel(traindata, alltestdata, "Prep.dur", "PREPpred"),         #PREP = constant
              LMEmodel(traindata, alltestdata, traindata$Anes.dur, "ANESpred"),       #ANES = LME
              LMEmodel(traindata, alltestdata, traindata$Proc.Duration, "PROCpred"), #PROC = LME
              constantModel(traindata, alltestdata, "Wrapup.dur", "WRAPpred")) %>%   #WRAP = constant
  transmute(Case. = Case.,
            LMEpart = PREPpred + ANESpred + PROCpred + WRAPpred)

#use RFmodel function to generate predictions
predictionsRF2 <- RFmodel(traindata = RFtraindata,                #train with original training data (=same mode
l)
                          testdata = select(newRFdata,-Case.),    #test with newRFdata, factors fixed (=new pred)
                          testcasenum = select(newRFdata,Case.))  #use Case# for newRFdata


#COMBINE PREDICTIONS
predictions2 <- list(alldata[c("Case.","Case.Date","Primary.Procedure.Description","Case.Duration","surgEst")],
                     predictionsSMA2,
                     predictionsLMEcase2,
                     predictionsLMEpart2,
                     predictionsRF2) %>% reduce(left_join, by = "Case.")
predictions2 <- left_join(newcasedata["Case."],predictions2, by = "Case.") #only keep new data predictions
```

Figure G-23 Round 2: Predictions

APPENDIX REFERENCES

[1] Mind Tools, "Frederick Taylor and Scientific Management," 2016. [Online]. Available: https://www.mindtools.com/pages/article/newTMM_Taylor.htm.

[2] Six Sigma Daily, "Henry Ford and the Roots of Lean Manufacturing," December 2017. [Online]. Available: https://www.sixsigmadaily.com/henry-ford-lean-manufacturing/.

[3] T. Hessing, "Walter A. Shewhart," SixSigmaStudyGuide.com, 2019. [Online]. Available: https://sixsigmastudyguide.com/shewhart/.

[4] Naomi, "W. Edwards Deming," SixSigmaStudyGuide.com, 2016. [Online]. Available: https://sixsigmastudyguide.com/william-edwards-deming/.

[5] E. Swan, "Grand-Daddy of Quality: Dr. W. Edwards Deming," Go Lean Six Sigma, October 2019. [Online]. Available: https://goleansixsigma.com/dr-w-edwards-deming-grand-daddy-quality/.

[6] T. Hessing, "Juran, Joseph M.," SixSigmaStudyGuide.com, 2017. [Online]. Available: https://sixsigmastudyguide.com/juran/.

[7] J. L. Smith, "Management: The Lasting Legacy of the Modern Quality Giants," Quality Magazine, October 2011. [Online]. Available: https://www.qualitymag.com/articles/88493-management--the-lasting-legacy-of-the-modern-quality-giants.

[8] Naomi, "Genichi Taguchi," SixSigmaStudyGuide.com, 2017. [Online]. Available: https://sixsigmastudyguide.com/genichi-taguchi/.

[9] ReliaSoft Corporation, "Taguchi Robust Design for Product Improvement," Reliability HotWire, April 2011. [Online]. Available: https://www.weibull.com/hotwire/issue122/hottopics122.htm.

[10] British Library, "Philip Crosby: A Zero Defects Thinker," 2015. [Online]. Available: https://www.bl.uk/people/philip-crosby.

[11] Strategos, "A History of Lean Manufacturing," February 2016. [Online]. Available: http://www.strategosinc.com/just_in_time.htm.

[12] R. Coetzee, K. van der Merwe and L. van Dyk, "Lean Implementation Strategies: How Are the Toyota Way Principles Addressed?," *South African Journal of Industrial Engineering,* vol. 27, no. 3, p. https://pdfs.semanticscholar.org/bdc7/6c915465a5cc0b21d4bc4a18943decd961c3.pdf, November 2016.

[13] Kanbanchi + G Suite, "What is Kaizen?," May 2015. [Online]. Available: https://www.kanbanchi.com/what-is-kaizen.

[14] J. P. Womack, D. T. Jones and D. Roos, The Machine That Changed the World, New York, NY: Simon & Schuster, 2007.

[15] J. P. Womack and D. T. Jones, Lean Thinking: Banish Waste and Create Wealth in Your Corporation, New York, NY: Simon & Schuster, 1996.

[16] B. Emiliani, "Is Lean the Same as TPS?," April 2016. [Online]. Available: https://bobemiliani.com/is-lean-the-same-as-tps/.

[17] Shmula.com, "Six Sigma DMAIC Training Table of Contents," 2020. [Online]. Available: https://www.shmula.com/six-sigma-dmaic-training-table-of-contents/.

[18] H. S. Gitlow and D. M. Levine, Six Sigma for Green Belts and Champions : Foundations, DMAIC, Tools, Cases, and Certification, Pearson Education, Inc. , 2005.

[19] B. Wheat, C. Mills and M. Carnell, Leaning into Six Sigma: The Path to Integration of Lean Enterprise and Six Sigma, Boulder City, Nev.: Pub. Partners, 2001.

[20] C. Intrieri, "What is Lean Six Sigma: A Combined Management Approach," Cerasis, October 2013. [Online]. Available: https://cerasis.com/what-is-lean-six-sigma/.

[21] Z. ShahabiKargar, S. Khanna, N. Good, A. Sattar, J. Lind and J. O'Dwyer, "Predicting Procedure Duration to Improve Scheduling of Elective Surgery," *PRICAI: Trends in Artificial Intelligence, Lecture Notes in Computer Science,* vol. LNCS vol 8862, pp. 998-1009, 2014.

[22] N. Hosseini, C. Jankowski and K. Pasupathy, "Surgical Duration Estimation via Data Mining and Predictive Modeling: A Case Study," *AMIA Annual Symposium Proceedings,* vol. 2015, pp. 640-648, 2015.

[23] J. J. Pandit, "Rational planning of operating lists: a prospective comparison of ‘booking to the mean’ vs. ‘probabilistic case scheduling’ in urology," *Anesthesia,* vol. 75, pp. 642-647, 2020.

[24] A. Larsson, "The accuracy of surgery time estimations," *Production Planning & Control,* vol. 24, pp. 891-902, 2013.

[25] R. K. Shukla, J. S. Ketcham and Y. A. Ozcan, "Comparison of Subjective versus Data Base Approaches for Improving Efficiency of Operating Room Scheduling," *Health Services Management Research,* vol. 3, no. 2, pp. 74-81, 1990.

[26] P. Stepaniak, C. Heij and G. De Vries, "Modeling and prediction of surgical procedure times," *Statistica Neerlandica,* vol. 64, no. 1, p. 1–18, 2010.

[27] F. Dexter and J. Ledolter, "Bayesian prediction bounds and comparisons of operating room times even for procedures with few or no historic data," *Anesthesiology,* vol. 103, no. 6, p. 1259–1267, 2005.

[28] F. Dexter, J. Ledolter, V. Tiwari and R. Epstein, "Value of a scheduled duration quantified in terms of equivalent numbers of historical cases," *Anesthesia & Analgesia,* vol. 117, no. 1, p. 205– 210, 2013.

[29] Y. Jiao, A. Sharma, A. B. Abdallah, T. M. Maddox and T. Kannampallil, "Probabilistic forecasting of surgical case duration using machine learning: model development and validation," *Journal of the American Medical Informatics Association,* vol. 27, no. 12, p. 1885–1893, 2020.

[30] C. Combes, N. Meskens, C. Rivat and J. Vandamme, "Using a KDD process to forecast the duration of surgery," *International Journal of Production Economics,* vol. 112, no. 1, p. 279–293, 2008.

[31] S. Devi, K. Rao and S. Sangeetha, "Prediction of surgery times and scheduling of operation theaters in optholmology department," *Journal of Medical Systems,* vol. 36, no. 2, p. 415–430, 2012.

[32] M. P. Davila, "A Methodology for Scheduling Operating Rooms Under Uncertainty," Scholar Commons: Graduate Theses and Dissertations, University of South Florida, 2013.

[33] P. Joustra, R. Meester and H. van Ophem, "Can statisticians beat surgeons at the planning of operations?," *Empir Econ,* vol. 44, pp. 1697-1718, 2013.

[34] J. H. May, D. P. Strum and L. G. Vargas, "Fitting the Lognormal Distribution to Surgical Procedure Times," *Decision Sciences,* vol. 31, no. 1, pp. 129-148, 2000.

[35] Z. Shahabikargar, S. Khanna, A. Sattar and J. Lind, "Improved Prediction of Procedure Duration for Elective Surgery," *Integrating and Connecting Care,* vol. 2, pp. 133-138, 2017.

[36] I. Wright, C. Kooperberg, B. Bonar and G. Bashein, "Statistical modeling to predict elective surgery time: Comparison with a computer scheduling system and surgeon-provided estimates," *Anesthesiology,* vol. 85, no. 6, p. 1235–1245, 1996.

[37] E. Kayıs, T. T. Khaniyev, J. Suermondt and K. Sylvester, "A robust estimation model for surgery durations with temporal, operational, and surgery team effects," *Health Care Management Science,* vol. 18, pp. 222-233, 2015.

[38] E. Edelman, S. van Kuijk, A. Hamaekers, M. de Korte, G. van Merode and W. Buhre, "Improving the Prediction of Total Surgical Procedure Time Using Linear Regression Modeling," *Frontiers in Medicine,* vol. 4, no. 85, pp. 1-5, 2017.

[39] J. Wang, J. Cabrera, K.-L. Tsui, H. Guo, M. Bakker and J. B. Kostis, "Predicting Surgery Duration from a New Perspective: Evaluation from a Database on Thoracic Surgery," 2017. [Online]. Available: http://arxiv.org/abs/1712.07809.

[40] M. A. Bartek, R. C. Saxena, S. Solomon, C. T. Fong, L. D. Behara, R. Venigandla, K. Velagapudi, J. D. Lang and B. G. Nair, "Improving Operating Room Efficiency:

Machine Learning Approach to Predict Case-Time Duration," *Efficient Operating Room and Machine Learning,* vol. 229, no. 4, pp. 346-354, 2019.

[41] K. W. Soh, J. Wallace, M. O'Sullivan, C. Walker and D. Grayson, "Case study of the prediction of elective surgery durations in a New Zealand teaching hospital," *Int J Health Plann Mgmt,* vol. 35, pp. 1593-1605, 2020.

[42] M. J. C. Eijkemans, M. van Houdenhoven, T. Nguyen, E. Boersma, E. W. Steyerberg and G. Kazemier, "Predicting the Unpredictable: A New Prediction Model for Operating Room Times Using Individual Characteristics and the Surgeon's Estimate," *Anesthesiology,* vol. 112, pp. 41-49, 2010.

[43] UCLA Statistical Consulting Group, "Introduction to Linear Mixed Models," UCLA Institute for Digital Research & Education, [Online]. Available: http://stats.idre.ucla.edu/other/mult-pkg/introduction-to-linear-mixed-models/. [Accessed 2021].

[44] G. Olbricht, "Lecture 34: Fixed vs. Random Effects," Purdue University, 04 2011. [Online]. Available: https://www.stat.purdue.edu/~ghobbs/STAT_512/Lecture_Notes/ANOVA/Topic_34.pdf. [Accessed 2021].

[45] G. K. Hajduk, "Introduction to Linear Mixed Models," Coding Club, 10 09 2019. [Online]. Available: http://ourcodingclub.github.io/tutorials/mixed-models/. [Accessed 2021].

[46] B. McGill, "Is it a fixed or random effect?," Dynamic Ecology, 04 11 2015. [Online]. Available: https://dynamicecology.wordpress.com/2015/11/04/is-it-a-fixed-or-random-effect/. [Accessed 2021].

[47] B. Winter, "A Very Basic Tutorial for Performing Linear Mixed Effects Analyses: Tutorial 2," University of California, Merced, Cognitive and Information Sciences, 19 05 2014. [Online]. Available: http://www.bodowinter.com/uploads/1/2/9/3/129362560/bw_lme_tutorial2.pdf. [Accessed 2021].

[48] A. Chakure, "Random Forest Regression," Medium, 29 06 2019. [Online]. Available: https://medium.com/swlh/random-forest-and-its-implementation-71824ced454f. [Accessed 2021].

[49] D. Mwiti, "Random Forest Regression: When Does It Fail and Why," neptune.ai, 05 2020. [Online]. Available: https://neptune.ai/blog/random-forest-regression-when-does-it-fail-and-why. [Accessed 2021].

[50] K. Powezka, P. Normahani, N. J. Standfield and U. Jaffer, "A novel team Familiarity Score for operating teams is a predictor of length of a procedure: A retrospective Bayesian analysis," *Journal of Vascular Surgery,* vol. 71, no. 3, pp. 959-966, 2020.

[51] S. Henrickson Parker, X. Lei, S. Fitzgibbons, T. Metzger, S. Safford and S. Kaplan, "The Impact of Surgical Team Familiarity on Length of Procedure and Length of Stay: Inconsistent Relationships Across Procedures, Team Members, and Sites," *World Journal of Surgery,* vol. 4, pp. 3658-3667, 2020.

[52] {R Core Team}, "R: A Language and Environment for Statistical Computing," 2021.

[53] H. Wickham, R. François, L. Henry and K. Müller, "dplyr: A Grammar of Data Manipulation," 2021.

[54] A. Zeileis and G. Grothendieck, "zoo: S3 Infrastructure for Regular and Irregular Time Series," *Journal of Statistical Software,* vol. 14, no. 6, pp. 1-27, 2005.

[55] D. Bates, M. Maechler, B. Bolker and S. Walker, "Fitting Linear Mixed-Effects Models Using {lme4}," *Journal of Statistical Software,* vol. 67, no. 1, pp. 1-48, 2015.

[56] H. Wickham, "forcats: Tools for Working with Categorical Variables (Factors)," R package version 0.5.1, 2021. [Online]. Available: https://CRAN.R-project.org/package=forcats.

[57] M. W. Andy Liaw, "Classification and Regression by randomForest," *R News,* vol. 2, no. 3, pp. 18-22, 2002.