Diagnosing Intensive Care Units and Hyperplane Cutting for Design of Optimal Production Systems

by

J Adam Traina

B.A. Harvard University, Cambridge, 2008
M.S. Harvard University, Cambridge 2008

Submitted to the Department of Mechanical Engineering and the MIT Sloan School of Management in partial fulfillment of the requirements for the degrees of

Master of Science in Mechanical Engineering

and

Master of Business Administration

in conjunction with the Leaders for Global Operations Program at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2015

Signature redacted

Author ....
Department of Mechanical Engineering and the MIT Sloan School of Management
May 8, 2015

Signature redacted

Certified by ....
Retsef Levi, Thesis Supervisor
J. Spencer Standish Professor of Management, Professor of Operations Management

Signature redacted

Certified by ....
Stanley Gershwin, Thesis Supervisor
Senior Research Scientist, MIT Department of Mechanical Engineering

Signature redacted

Approved by ....
David Hardt
Chair, Mechanical Graduate Program Committee

Signature redacted

Approved by ....
Maura Herson
Director, MBA Program, MIT Sloan School of Management

Diagnosing Intensive Care Units and Hyperplane Cutting for Design of Optimal Production Systems

by

J Adam Traina

Submitted to the Department of Mechanical Engineering and the MIT Sloan School of Management on May 8, 2015, in partial fulfillment of the requirements for the degrees of
Master of Science in Mechanical Engineering
and
Master of Business Administration

Abstract

This thesis provides a new framework for understanding how conditions, people, and environments of the Intensive Care Unit (ICU) effect the likelihood the preventable harm will happen to a patient in the ICU. Two years of electronic medical records from seven adult ICUs totalling 77 beds at Beth Israel Deaconess Medical Center (BIDMC) were analysed.

Our approach is based on several new ideas. First, instead of measuring safety through frequency measurement of a few relatively rare harms, we leverage electronic databases in the hospital to measure *Total Burden of Harm*, which is an aggregated measure of a broad range of harms. We believe that this measure better reflects the true level of harm occurring in Intensive Care Units and also provides hope for more statistical power to understand underlying contributors to harm.

Second, instead of analysing root causes of specific harms or risk factors of individual patients, we focus on what we call *Risk Drivers*, which are conditions of the ICU system, people (staff, patients, families) and environments that affect the likelihood of harms to occur, and potentially their outcomes. The underlying premise is that there is a relatively small number of risk drivers which are common to many harms. Moreover, our hope is that the analysis will lead to system level interventions that are not necessarily aiming at a specific harm, but change the quality and safety of the system.

Third, using two years of data that includes measurements of harms and drivers values of each shift and each of seven ICUs at BIDMC, we develop an innovative statistical approach that identifies important drivers and High and Low *Risky States.* Risky States are defined through specific combinations of values of Risk Drivers. They define environmental characteristics of ICUs and shifts that are correlated with higher or lower risk level of harms.

To develop a measurable set of Risk Drivers, a survey of current ICU quality metrics was conducted and augmented with the clinical experience of senior critical care providers at BIDMC. A robust machine learning algorithm with a series of validation techniques was developed to determine the importance of and interactions between multiple quality metrics. We believe that the method is adaptable to different hospital environments.

Sixteen statistically significant Risky States ($p < .02$) where identified at BIDMC. The harm rates in the Risky States range over a factor of 10, with high risk states comprising more that 13.9% of the total operational time in the ICU, and low risk states comprise 38% of total operating shifts.

The new methodology and validation technique was developed with the goal of providing a basic tools which are adaptable to different hospitals. The algorithm described within serves as the foundation for software under development by Aptima Human Engineering and the VA Hospital network with the goal of validation and implementation in over 150 hospitals.

In the second part of this thesis, a new heuristic is developed to facilitate the optimal design of stochastic manufacturing systems. The heuristic converges to optimal, or near optimal results in all test cases in a reasonable length of time. The heuristic allows production system designers to better understand the balance between operating costs, inventory costs, and reliability.

Thesis Supervisor: Retsef Levi
Title: J. Spencer Standish Professor of Management, Professor of Operations Management

Thesis Supervisor: Stanley Gershwin
Title: Senior Research Scientist, MIT Department of Mechanical Engineering

Acknowledgments

There are many, many people without whom the completion of this work would not have been possible. My journey through MIT has been at times overwhelming, humbling and elating yet it is the support from my friends, family and advisers during the challenging times that I most appreciate and recognize as keys to what success we have had in our journey of discovery together.

To my friends and family, thank you first and foremost. You are the pillar which stands in the face of all obstacles. Through the long journey of graduate school you have made the highs higher and the lows tolerable, turning even the darkest moments into opportunities to grow.

Retsef Levi has found a balance between giving my research creativity room to thrive with focus and academic rigour necessary to convince a scientific audience that these new ideas have tremendous merit. His keen interest in total project management, from interoffice relationships to academic rigor have helped me develop a multitude of tools that I will take with me into the future. Dr. Pat Folcarelli and Dr. Daniel Talmor's steadfast support and guidance for my investigation through the broad and unknown wilderness of medical care was a guiding light that I could always count on for support. Agnes Hu, my constant companion through this journey, I have grown to appreciate your drive for attention to detail and a deep respect for your unrelenting drive for understanding and truth, in spite of all obstacles. Yi Yin Ma, I thank you for your thoughtfulness and dedication to our work together. I am also grateful for the empathy, insight and enthusiasm of the nursing staff at BIDMC. Veronica Kelly, Lynn Mackinson and Juliann Corey your curiosity and helpfulness at all times throughout this research made working with you a tremendous pleasure.

Dr. Stanley Gershwin, to whom I owe the most gratitude, your contributions to my education and completion of this thesis I will never be able to thank you enough. Your patience over our last 4 years together has been instrumental as I adapted to MIT. You supported my growth in the directions that showed how much you care and your genuine desire to help. My second crack at research with BIDMC went farther and more smoothly than I ever expected, and I believe it is because you did all the hard work of training me to think, to experiment and to write. You supported my investigative curiosity, and through your tremendous effort we crafted a raw idea into a sensible contribution. Your tenacity, humor and generosity are limitless.

Finally, I feel it is appropriate to sadly acknowledge the loss of a dear friend who I once imagined would be graduating with me at the end of a long journey that we started together at MIT. Your brilliance and intensity burned brighter than anyone else I have ever known, and your dedication to others was an inspiration that will never grow dull.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis is presented in two unrelated parts. Chapters 1-5 are focused on developing a new risk management methodology to understand and control adverse events in healthcare environments. Chapter 6 presents an efficient algorithm to help manufacturing companies design profitable production systems.

1.1 Company Background

Deaconess Hospital was founded in 1896 by the Methodist Deaconesses to care for the Boston residents. The 14-bed infirmary was run by religious women who were passionate about caring for the sick and the poor of "every creed and race and social condition." In 1916, Beth Israel Hospital was founded by the Jewish community in response to the growing population of immigrants in Boston. With 45 beds, Beth Israel served kosher food and conducted religious services according to the Jewish faith. Like the Deaconess, Beth Israel offered, "medical and surgical aid and nursing to sick or disabled persons of any creed or nationality."

After tremendous growth, historic research accomplishments and partnerships with the Harvard Medical School, Beth Israel and Deaconess Hospital merged in 1996 to form the Beth Israel Deaconess Medical Center (BIDMC).

In 2011, BIDMC launched The Center for Healthcare Delivery Science. The mission of the center was to ensure an " integrated approach to understanding healthcare systems." The Center for healthcare Delivery Science is independent of any single academic department and supports cross functional disciplines to better understand healthcare delivery. Projects have spanned anaesthesia, critical care, gastroenterology, nephrology, nursing, orthopaedics, pharmacy, primary care, psychiatry, social work, and surgery. Progression towards higher reliability is demonstrable by both internal and external metrics. Internally, projects have decreased harms such as ventilator-associated pneumonia, bloodstream infection, and surgical site infection. External recognition for BIDMC's quality improvement efforts include the Truven Healthcare " top 100 hospitals," which consistently recognized BIDMC as one of 15 top academic medical centers nationally. BIDMC has also earned an "A" in Leapfrog's first-ever hospital safety grades and was selected as one of their top 65 hospitals in the country in 2011—the fifth time in six years. Additionally, BIDMC was awarded the Society of Critical Care medicine's 2010 Family-Centered Care Award, which recognizes innovation to improve the care provided to critically ill patients and their families and is given to only one hospital in the country each year. Lastly, in 2013, BIDMC was selected for the American Hospital Association-McKesson "Quest for Quality" award, which is given to the one hospital nationally that is best able to demonstrate progress among all six dimensions of quality as defined by the Institute of Medicine.

Today BIDMC sees nearly three quarters of a million patients annually at its Boston and satellite hospital locations. In addition to the main campus in the Longwood Medical Area of Boston, BIDMC has hospitals in Milton, Needham and Plymouth, outpatient clinics in Boston, Chestnut Hill, Lexington and Chelsea and is affiliated with health centers in Dorchester, Roxbury, Allston, Quincy and other areas.

1.2 Gordon and Betty Moore Foundation

The Gordon and Betty Moore Foundation is a private philanthropic organization established by Intel co-founder Gordon and his wife Better to "create positive change for future generations." The foundation establishes strategies and partners to fund "lasting, meaningful change" in environmental conservation, patient care and science. The patient care programs imagine a world, "where medical harms no longer occur. Where technology is connected and systems talk to each other. Where doctors, nurses, patients and families work as a team and decisions about health care are shared. Imagine a health care system that is so finely tuned it can eliminate preventable harms, cut health care costs and give patients and families the voice and care they deserve."

In September 2013, BIDMC was awarded \$ 5.3 million dollars through the Gordon and Betty Moore Foundation for their proposal titled " Optimizing ICU Safety through Patient Engagement, System Science and Information Technology." Three work streams are supported by the Foundation, evolving into eight outcomes. This project falls under the purview of the first workstream, " Managing Risky States to Prevent Harm," and is concerned with identifying risky states and leading harm indicators.

As part of the Grant, BIDMC became a member of the Libretto ICU Consortium. The Libretto ICU Consortium includes the Beth Israel Deaconess Medical Center, Brigham and Women's Hospital, Johns Hopkins University Hospital and University of San Francisco Medical Center. This collaborative was formed to support efforts which aim to eliminate preventable harms and unnecessary costs by engaging patients and families in their own care within a re-designed, supportive healthcare system.

Funded by the Gordon and Betty Moore Foundation, each center is collaborating on solutions that will be scalable and spreadable to ICU patients in all hospitals. While each member of the collaborative is developing their own approach, a measurement and evaluation representative from each center has worked together over the past year to work to define standard measurement definitions for seven preventable harms that ICU patients experience (Table 1.1). Other collaborative groups are working collectively on methods to increase patient and family engagement and to improve end of life discussions and decision making.

Beth Israel Deaconess Medical Center's long-term goal is to eliminate preventable harm in healthcare. Working with the Libretto Consortium represents a critical milestone in that journey by addressing three fundamental, interrelated barriers to the elimination of preventable patient harm on a broader scale: (1) unreliable systems of care, coupled with a lack of technical expertise

	Harm
a.	Central Line Associated Bloodstream Infections (CLABSI)
b.	Iatrogenic harm from ventilators
c.	Ventilator associated events (including pneumonia and avoidable patient days on ventilator)
d.	Deep Venous Thrombosis (DVT)-Pulmonary Embolism (PE)
e.	ICU-Acquired delirium and weakness
f.	Loss or diminution of respect and dignity afforded to patients and families
g.	Inappropriate care and excessive intensity of care

Table 1.1: Harms Standardized by The Libretto Consortium

as to how to improve those systems; (2) failure to adequately engage patients and families in their own care; and (3) failure to spread successful innovation into non-academic hospital settings.

1.3 Brief History of Healthcare Safety and Quality

One of the earliest documented breakthroughs in the improvement of healthcare delivery was the adoption of washing hands between patients. At the General Hospital of Vienna in 1846, Ignaz Semmelweis observed there was a stark difference in fever rates among pregnant women between the First Clinic and the Second Clinic [Semmelweis 1983]. The first clinic was staffed by students and physicians, whereas the Second Clinic was run by midwives. He noticed that physicians who went directly from the autopsy suite to the obstetrics ward had a "disagreeable odor" on their hands, even after washing their hands with soap and water before they entered the obstetrics clinic. He postulated that the puerperal fever was caused by "cadaverous particles" transmitted from the autopsy suite to the obstetrics ward via the hands of students and physicians. In May 1847, he insisted that students and physicians clean their hands with a chlorine solution between each patient in the clinic. The maternal mortality rate in the First Clinic dropped and remained low for years. This represented the first evidence indicating that an operational change to the way that healthcare providers did their regular work, in this case hand washing with an antiseptic agent between patient contacts, could improve the quality of healthcare delivery by reducing the transmission of contagious diseases.

More recently, the Institute of Medicine, a government funded organization dedicated to improving healthcare quality, published a 1999 study entitled To Err is Human: Building a Safer Health System, which estimated that over 98,000 patients in the US die annually due to preventable medical errors. In 2001 the same institute published a follow-up report called Crossing the Quality Chasm: A New Health System for the 21st Century which detailed a nationwide shortcoming to provide quality healthcare. In particular, they claim that US hospitals lack " ...the environment, the processes, and the capabilities needed to ensure that services are safe, effective, patient-centered, timely, efficient, and equitable."

In the early 2000s, healthcare organizations began to measure healthcare quality, leading to major improvements. In 2002 The Joint Commission on Accreditation of Hospitals developed nationally standardized quality measures for conditions such as heart failure, acute myocardial infarction, pneumonia and pregnancy. The commission required all accredited hospitals to begin reporting metrics on these standards, and they began publishing the data in 2004. Also in 2004, the Centers for Medicare and Medicaid Services instituted a financial incentive program to reward and penalize hospitals that did not report their metrics. The Joint Commission developed six national measures of quality for ICUs, but implementation efforts were suspended before they began in 2005 due to an organization move to refocus efforts on surgical operations. In the second half of the last decade, the Joint Commission and Center for Medicare and Medicaid Services have expanded their reporting requirements and publications for public consumption.

In 2009, studies found significant improvements in the measure of healthcare quality that were enforced by the Joint Commission. For example, in 2002, 87.3% of eligible patients with acute myocardial infarction were discharged with beta blockers. In 2009 that metric rose to 98.3%.

While many of the Joint Commission's metrics accurately measure activities that improve patient outcome, not all of their proposed metrics were easy for hospitals to comply with and not all of the measures led to better patient care. In 2010 the Joint Commission published a revised set of metrics with the goal of maximizing the likelihood of improved patient care while minimizing unintended complications and non-value added work for healthcare providers. However, once again these metrics have not been implemented in a meaningful way. In fact, a recent multicenter study showed that there is no comprehensive benchmark set of metrics for healthcare quality in critical care [Martinez et al., 2013].

1.4 Project Background

The Intensive Care Unit (ICU) is a complex healthcare environment that provides advanced life support for critically ill patients. It is estimated that patients experience 1.7 medical errors per day in the ICU [Pronovost et al., 2005]. BIDMC has 7 adult intensive care units with a total of 77 patient beds. Each ICU specializes in an area of medicine, as shown in the table below, and is equipped to provide support for patients from other ICUs. Across the hospital, BIDMC records over 700 adverse events or near misses every year in it's voluntary reporting system, which is expected to vastly underreport incident rates [Taylor et al., 2004].

Unit	Abbr.	Specialty	Capacity
Medical ICU	MICU	Medical Patients	16
Surgical ICU	SICU	Surgical Patients	7
Trauma Surgical ICU	TSICU	Trauma and Surgical Patients	13
Cardiovascular ICU	CVICU	Cardiovascular patients	14
Finard 4	Fin4	Medical, General Critical Care	13
Coronary Care Unit	CCU	Heart Patients	7

Table 1.2: Critical Care Wards at BIDMC

Preventable harm in the critical care environment is a contested concept with several authorities in the field considering a veirty of methods to identify and measure these concepts [Pronovost et al., 2005, Classen et al., 2011]. The Libretto ICU Consortium is currently defining a narrow set of preventable harms that are common across institutions. Later sections in this thesis describe different events and preventable harm that are measured at BIDMC's ICUs (See Table 3.1).

This project is the central component for three deliverables from the Moore Grant with the objective of providing tools to address the issues of reducing preventable harm in critical care:

Output 1: Measurement Strategy including a definition of the burden of harm in the ICU and a measurable, meaningful decrease in the burden of harm.

Output 2: Managing Risky States to Prevent Harm with a model for measuring and managing the leading harm indicators that create "risky states," including a dashboard alert system coupled with a mitigation approach.

Output 3: Spread of these strategies and models to Community Hospital Partners, including IT applications.

1.4.1 Outline

This thesis will present a measurement strategy and model to show that measures of the conditions of the ICU, the people in the ICU and the environment of the ICU that affect the likelihood of harms and their magnitude can be used to suggest working environments (states) that might lead to greater frequency of preventable harm. Nearly 38% of ICU shifts will be classified as low risk states which are significantly below the mean. About 13% of shifts will be classified as significantly high risk states.

Preventable harm and near misses are more than ten times more likely to occur than in the low risk states. It will be shown that high risk states are characterized by **combinations** of elevated metrics related to workload, acuity, and unfamiliarity with tasks, as well as other measures.

Low risk states will be shown to correlate with lower values of the same metrics, and that risk states remain low during the elevation of a single measure.

Chapter 2 provides a review of prior work on safety and risk management at the BIDMC and other hospitals; prior work has focused primarily on single interventions for specific adverse events. A new approach to risk management and safety is proposed which considers aggregated metrics for ICU wards and their impact on multiple harms. s.

Chapter 3 describes the methodology for aggregating and measuring preventable harm. It goes on to show how the Intensive Care Unit environment, people and processes can be quantified and measured. This measurement strategy sets the foundation to design system level interventions to reduce the likelihood of many adverse event

Chapter 4 describes the mathematical models used to identify the prevalence of preventable harm in different groups of shifts. Validation techniques to confirm that the mathematical models are stable and accurate.

Chapter 5 presents the results of the modelling methodology. It is shown that combinations of adverse environments, such as high workload, new nurses and high numbers of overflow patients, are indicators for increased exposure to adverse events. It is also shown that the ICU is relatively safe when only one indicator for high risk in present.

Chapter 6 discusses these observations and suggests directions for future research.

Chapter 7 presents unrelated work in mixed integer nonlinear programming. A heuristic is developed to help production system designers know when to invest in more reliable processes.

Chapter 2

Background and Preliminaries

This chapter begins with a review of traditional approaches to risk management and safety in hospitals and ICUs in particular, with a discussion of some of their weaknesses. We begin with a more general review of the psychological basis for cognitive errors and move to reviewing current risk management and safety practices in healthcare environments. We highlight the pros and cons of current practices, and in the final section, outline a new paradigm of environmental risk management to address some of the shortcomings of current methods.

2.1 Review of Human Errors

In spite of the rigorous training processes and highly qualified individuals that work in healthcare delivery teams, many errors continue to occur, particularly in the critical care environment. We delve into prior work on human errors because a basic understanding of why errors happen can foster new ways to understand, measure and mitigate errors.

Many psychologists have researched the basis for human errors, but Reason [1990] provides a unified framework that describes the main themes for cognitive errors. He proposes a foundational model for brain function, claiming that much of human daily activity is controlled automatically and effortlessly by the brain. A person does not consciously demand the flexion of each individual muscle when they take a step, or pour a glass of water. Instead, we carry a set of mental *schemata* that control each minute action that is repeated. These schemata are unconsciously operated for short periods of time by the brain and require few mental resources. Working to complement the *schematic control mode* is the *conscious part* of the brain called the *attention control mode*, which is used for problem solving and monitoring of the schematic control mode. The attention control mode is slow, sequential, draws a great deal of mental resources, and has limited capacity. In contract, the schematic mode is fast, effortless and operates in parallel.

Rasmussen and Jensen [1974] suggest three cognitive modes for human performance that connect to Reason's model for operating modes. The first mode is *skill-based operation*, a phase during which actions are governed by the schematic control mode and are largely preprogrammed. The second phase is *rules-based*, which combines aspects of schematic control and attention control; new problems are solved with rules that are similar or familiar to problems previously solved and stored in the unconscious system. The third mode is *knowledge-based*, in which novel situations require heavy reliance on the attention control mode to seek new solutions in unfamiliar environments.

Reason and Rasmussen classify errors for each of the three cognitive modes. *Skill based errors* are called "Slips," and refer to unconscious errors in automatic activity. Slips can occur from a lack of additional checks. *Rule based errors* occur when the problem solving area of the brain is active and the wrong rule or schema is chosen – often due to a misperception of the situation and subsequent misapplication of a rule, usually one that is frequently used and may seem to be a close substitute. This type of error can occur more frequently when the conscious part of the brain is overloaded, and can be influenced by practice. *Knowledge based errors* occur when an individual confronts a novel problem with which he or she has no prior experience. The pattern matching ability of the brain fails, and habits can alter matching or calculations and lead to mistakes. Flavours of these errors include biased memory (psychological biases that enhance or impair information recall), availability heuristic (over-reliance on immediate examples), confirmation bias (searching to confirm one's beliefs) and overconfidence (over-reliance on personal judgement) [Leape, 1994].

2.2 Traditional approaches to risk management and safety in healthcare

In general, there is no broadly adopted set of quality metrics for ICUs [Martinez et al., 2013]. As a result, many institutions measure the same harmful outcomes in different ways. For example, Berenholtz et al. [2002] finds that mortality rates in ICUs was a common measure for ICU quality. However, its widespread adoption as a standard of measure was quickly opposed, due to issues with co-morbidities and geographic region leading to "requirements for risk adjustment and burdensome data collection" that were not properly accounted for with the crude measure [Pronovost et al., 2001].

Additional debates over measurement standards and meaning have been waged over the meaning of ICU lengths of stay (LOS), average mechanical ventilation and patient satisfaction [Berenholtz et al., 2002]. However, broad adoption of these measurements failed largely due to the complexity of measuring adherence to specified procedures and the broad range of difference between hospital goals, patient expectations subsequent disagreements over the meaning of the metrics.

Patient measures:

Even though there are no global standards, a great deal of work has been done to measure harmful events in the ICU and to connect these single contributors to specific healthcare delivery errors on individual patients utilizing well-known techniques borrowed from systems engineering.

Most metrics have been developed to measure patient treatment and outcomes; De Vos et al. [2007] reviews over 62 different ICU metrics and identifies only 11 quality indicators that are relevant to Dutch hospitals, such as spikes and drops in individual glucose levels and the number of unplanned extubations.

Successful interventions at the level of improving a single process have been demonstrated widely, resulting from the application of lean, six sigma, and total quality concepts appear in over thirty recent publications which are summarized by Mazzocato et al. [2010]. Some of the tools to investigate sources of risk in these works include the 5 Whys, value stream mapping and failure modes and effects analysis, all of which follow the treatment path of one patient at a time. The corrective actions for each individual process that evolve from these investigations include 5S, Kanbans, and the specification of "Standard Procedures."

Interventions:

These tools have allowed for significant progress in the reduction of the frequency of specific healthcare delivery errors and harms. Perhaps one of the most successful examples is the use of check-lists in both the ICU and the Operating Room [Haynes et al., 2009]. In medicine, a check-list is a list that specifies processes steps which must be completed one after another during a medical intervention. Ideally the items on the list are checked off by a practitioner or a clinical team member as they are completed, ensuring that the specific tasks are finished. In an 18 month intervention study, check-lists that dictated standard work, process completion and inspection steps have been shown to reduce the rate of catheter related bloodstream infections in ICUs by 66% [Pronovost et al., 2006]. Other studies show that check-lists have improved attention to procedural adherence and clinical staff morale without impacting mortality rates or other ICU quality metrics [Simpson et al., 2007].

Check-lists seem to be effective in reducing the frequency of "slip" errors in the critical care environment because they remind a trained practitioner which patterns to follow in a given situation. However, they are not an effective tool for preventing rules-based errors because they do little to intervene in cases of misperception. While check-lists provide built-in double checks and can help eliminate some mistakes, the addition of the extra steps and the creation of a complex set of rules can actually increase errors; complex check-lists can increase errors particularly in situations where the care path is not fully predictable.

Additionally, these lean, or system engineering based tools, does a good job at targeting repeatable phenomena. Processes or interventions that recur in the same manner while holding all the other variables constant can be mapped and improved, allowing practitioners to shift from attention control mode to faster, more proficient and more robust actions in schema mode.

These tools do not match many of the observed harms in the Critical Care Environment at BIDMC, particularly when used on care processes in the ICU that are highly complex and unpredictable. There are approximately 700 patient safety events reported annually at BIDMC. While the majority of these events represent an intercepted risk to safety (i.e., "near miss") as opposed to actual patient harm, each of these events is considered an important indicator on the quality of care in the hospital.

These 700 safety events consist of a wide range of harm events actions, many of which have a very low probability of re-occurrence, but a very high impact should they occur. A root cause analysis (RCA) is conducted for each major adverse harm event that resulted in serious patient harm, and a detailed mitigation plan is developed to eliminate or at least reliably reduce the risk of another patient experiencing that specific type of harm. Typical mitigation strategies championed by BIDMC's Department for Health Care Quality and Patient Safety include additional training, automated alerts, and process standardization. However, the corrective actions developed after a given adverse event occurs do not necessarily offer an effective means to address other low probability, high impact risk events.

With the current system at BIDMC, measurements are focused on solving one problem at a time and are treated somewhat independently from the other events and systems that operate within the hospital. The reporting system is a lagging indicator, showing harm only after it occurs instead of warning about approaching issues. Due to the independence of each investigation, finding solutions can be costly in time and money, and may impact other hospital systems in unpredictable ways.

Lean, root cause analysis and check-list based tools work well for repeatable situations, but are not as efficient in helping practitioners with novel, rare, complex, and not fully predictable problems that challenge the conscious mind and help them solve new situations which they face on a daily basis.

Early Ward Measurements:

De Vos et al. [2007] considers ward level indicators, such as bed utilization, nurse to patient ratio, and intensivist availability. These system level workload indicators are potentially powerful, but when interpreted without context they can be mis-leading as due to the multitude of explanatory factors that may indicate elevated levels in these metrics but not an effect on harmful outcomes to patients. An example of this is the issue of readmission: Chen et al. [1998] suggests that the number of readmissions to the ICU within 48 hours of discharge is an indicator of ICU quality. However, Angus [1998] points out that readmission alone is not a sufficient metric, as it may be a sign of many clinical decisions, such calculated decisions to temporarily move a patient out of the ICU while the don't need critical care, managing with chronic conditions that require re-hospitalization, or taking the risk of an early discharge at the request of the patient or healthcare proxy.

2.3 New approach to healthcare safety and risk management

Managing preventable harm in the ICU with a systems engineering approach is a new problem in medicine with a sought after solution [Pronovost and Bo-Linn, 2012]. Successful paradigms for engineering system safety have recently been demonstrated in the airline and computer security industries [Leveson, 2011]. These approaches treat human errors as a symptom of poor system design and suggest problem solving architectures to measure system performance and conditions that influence reliable outcomes. Leveson suggests several motivations for new approaches to safety:

Fast Pace of Technological Change:

In the beginning of the 20th century, it took nearly thirty years to translate a basic technical discovery into a commercial product. Today technologies are commercialized in two to three years and may be obsolete in five. The rapid growth and introduction of new technology shrinks the opportunity to learn from past mistakes and increases the continuous training required to leverage new products.

Changing view of public safety:

In today's healthcare society, individuals have limited control over the risks to which they are exposed when seeking medical treatment. Consumers are demanding that healthcare institutions develop various forms of oversight and regulation to control rising costs and malpractice.

This thesis deals with the following challenges to system safety in healthcare delivery environments such as the ICU:

1. Specific adverse events are relatively rare, making them difficult to characterize and analyze statistically.

2. There are many contributing factors that may impact adverse harm events, obscuring causal analysis.

3. System studies can be costly and time consuming.

Our approach is based on several new ideas. In particular, instead of tracing individual harmful events and seeking to understand their respective causes, this thesis seeks to understand the environmental factors that increase the occurrence of multiple types of harms. Next we outline the major ideas underlying the approach taken in this thesis to overcome the three challenges above:

1. Different types of harm events can be measured and aggregated to create a notion of "Total Burden of Harm." The "Total Burden of Harm" refers to the totality of undesirable harm events that may or may not cause a tangible harmful outcome to a patient.

2. There are measurable environmental factors, called "Drivers," are common to many harms. Drivers are conditions of the ICU, the people in the ICU and the environment of the ICU that affect the likelihood of harm occurring and the magnitude of the respective outcome. Drivers are not specific to a particular incident or patient, but describe more generic conditions of the healthcare environment.

3. Statistical approaches could be applied to cluster observations of drivers around observations of harms to find ranges of driver values in which more or less harms occur. These discrete clusters in driver parameters space are collectively called "Risky States," and are defined though combinations of drivers and their respective range values.

To make these definitions more illustrative, consider briefly an example of a car which has an accident. The specific accident could be caused by the fact that the driver could not stop on time, lost control of the vehicle, or experience a mechanical failure. However, snow roads, fog, badly maintained car, lack of safety measurements, tired or incompetent driver are all examples of drivers what would increase the likelihood of a car accident and the magnitude of its potential outcome.

The Risky State of the car would be the range of specific conditions that are correlated with a probability of an accident; for example on roads with more than 1 inch of snow with a driver that has slept less than 6 hours in the last day, the chance of crashing would likely be higher than if the driver had slept for more than 6 hours.

Specific harms in the healthcare environment are relatively infrequent, so by aggregating individual harms together and treating them as a general undesirable phenomenon, one could hopefully apply powerful statistical tools to draw conclusions about common drivers that apply to all types of harm. The full definition of the Total Burden of Harm and all the harms that could be considered therein was developed by Hu [2015], and includes nearly 4,000 harmful events over two years in the ICUs at BIDMC. The harms defined by the Libretto ICU Consortium comprise only 2.5% of these events. The analysis in this thesis focuses on harms which are directly attributable to the same shift on which they occur, capturing 12% of the Total Burden of Harm at BIDMC.

To demonstrate how the risk driver framework described above could be applied in the ICU environment, a preliminary review of 20 serious adverse events reported from BIDMC ICUs was conducted. Each of these cases was reviewed in depth with the hospital team and Dr. Retsef Levi to identify risk drivers or conditions among the staff, ICU unit or organization as a whole that may have contributed to the event.

One example of a risky state revealed in this preliminary review is the fundamental mental condition of clinicians failing to recognize that they face a "special" scenario that falls outside of their routine practice. This risky condition could arise in various settings; for example, when a neurology patient is placed into a unit that typically does not routinely care for neurology patients, or when an interventional procedure is done in an ICU when that is not typically the location where the procedure is performed. The practitioner's unfamiliarity with the changes in their environment exposes them to *rules-based* errors.

The risk driver framework developed in this thesis provides a view of how drivers can affect multiple events and harms. This framework is a fundamental shift from the current approach to patient safety, which focuses on highly specific activities that are tightly linked to a single specific type of harm, such as check-list interventions for catheter-related bloodstream infections [Pronovost et al., 2006]. The results of this preliminary analysis have been used to develop an early model to identify high risk conditions in the ICU setting.

A pervasive challenge to this type of approach is the so-called curse of dimensionality; with so many influential factors driving healthcare delivery errors, it quickly becomes impossible to acquire enough data to create a model for even modest combinations of independent variables [Bellman et al., 1961] to identify causes for each individual type of harm.

Fortunately, recent advances in machine learning have simplified the task of determining which features of a problem are relevant and how to model them. Classification trees, originally developed in 1984 [Breiman et al., 1984] have been used with success to determine the salient features of multidimensional problems [Kira and Rendell, 1992]. Chapter 3 will explore the definitions and methods that were used to predict harm.

By leveraging dimensionality reduction and clustering techniques from machine learning, many factors can be considered simultaneously. Aggregating rare harmful events provides sufficient statistical to discover patterns within the healthcare environment may emerge which influence error rates.

Chapter 3

Measurement Methods

The goal of this chapter is to develop a descriptive framework for shifts and harms the ICU. The measurements in this chapter differ from other healthcare measures in that they are aggregated to the unit and shift level, instead of the patient level. Specifically, each shift and each ICU ward are described through a set of Risk Drivers (conditions) that can be thought of as the *independent variables* and an aggregated indicator of whether some harm event occurred in the respective shift or ward. In what follows, we describe the set of harms (discussed briefly in Section 3.1) and in more detail in Hu [2015] and the set of drivers that constitute our model.

Our model aims to describe the environmental conditions of the ICUs at BIDMC over time (shift by shift). The goal of the statistical modelling described in chapter 4 is to identify *High Risk* and *Low Risk* States. States are defined through a combination of drivers and their respective ranges of values. The statistical model aims to partition the driver space into states, whereas high/low risk state have higher/lower harm rates than the average overall shifts in the ICU.

To identify High and Low Risk States, we rely on aggregating individual patient data. This aggregated approach has several advantages:

1. Consideration of a larger spectrum of harm.

   Many ICU quality efforts consider only one healthcare delivery harm at a time [Pronovost et al., 2006, Mazzocato et al., 2010, De Vos et al., 2007]. Consideration of multiple harms may lead to a different understanding of the environment when are correlated with preventable harm.

2. Improved statistical power.

   The aggregation of harm allows the use of more powerful statistical techniques. Rare events that characterize individual harms at BIDMC can be difficult to predict. By aggregating them together into a "Total Burden of Harm," statisical power is improveed and conclusions can be drawn about undesirable harms in general.

3. Foundational work to develop leading indicators of harm.

   Classification of hospital shifts as "High Risk" or "Low Risk" based on retrospective analysis lends insight into the level of safe operating conditions on those shifts. If it becomes possible to estimate the majority of events on future shifts, levels of risk can potentially be predicted and mitigated before harms occur.

4. Interventions may be systematic and effect the reduction of multiple types of harm.

   Since measurements are made at the unit level and conclusions may be drown about multiple types of harm, an intervention that alters the state of a shift may reduce the likelihood of harm occurring.

One of the primary technical challenges to a holistic approach to healthcare system safety is the wide variety of reported harm incidences and scarcity of repetition of each particular type of harm. Investigating the chain of events that led to each of the 700 harms in BIDMC's reporting system was intractable. Instead, we use the "Total Burden of Harm," which reflects the aggregation of different types of harms. The Total Burden of Harm allows for enough repeatable observations of harm to use powerful statistical tools. These tools help us understand the influences that a single driver, or combination of drivers, has on multiple types of harm. We seek to understand the environmental drivers that contribute to multiple types of harm, with the hope that interventions in these drivers, such as redistribution of workloads, better mixes of staff experience or new technologies, will reduce the Total Burden of Harm.

The definitions of specific harms and drivers presented in this section were developed and refined over a year of cross functional workshops that brought together senior staff in critical care.

Physicians, nurses, nurse managers, department chiefs, data scientists, statisticians and engineers worked together to determine environmental drivers and harms that could be measured and had meaningful definitions. Disparate databases were mined and combined from hospital departments, including electronic medical records, billing, staffing and utilization monitoring systems, spanning two years from 2012 through 2013. Over 10,000 critical care shifts in six adult Intensive Care Units were analyzed. The wards included the MICU, SICU, TSICU, CCU and Finard 4. The CVICU was eliminated from the dataset as an outlier because its use is combined as a surgical recovery room.

3.1 Definition of Harm in Healthcare Delivery

We rely on the work of Hu 2015 for definitions of the Total Burden of Harm in the ICUs at BIDMC. Table 3.1 summarizes the harms considered in this thesis. These harms were identified though three primary sources. First, consultation with the institute for Health Improvement's Trigger tool [Griffin and Resar, 2009] led to the development of an innovative automated large scale search though the medical IT recods at BIDMC to identify harms. Second, representatives from the Libretto ICU Consortium aided the development of the definitions and measurements of other harms. Finally, additional harms were developed based on the clinical expertise of the senior nurses and physicians in the ICUs at BIDMC and a review of over thirty harms from the BIDMC incident reporting system.

The harms in Table 3.1, which are considered in the remainder of this thesis, represent the subset of harms which occur *instantaneously* on a given shift. For example, a patient who falls in the ICU is considered an instantaneous harm because the harm event clearly occur on that shift. In contrast, a ventilator associated infection may develop over several shifts, and its precise time of occurrence can be difficult to ascertain. The full definition of the Total Burden of Harm was developed by Agnes Hu, 2015 and includes nearly 4,000 events over two years in the ICUs at BIDMC. The analysis in this thesis is focused on harms which are instantaneous and capture 12% of the Total Burden of Harm. About 3.5% of shifts over the two year data set have at least one of these instantaneous harms. Each shift that is associated with the value 1 if a harm occurred and 0 otherwise.

Harm	Definition
Arrest	Cardiac Arrest or Code Blue
Code Purple	Shifts on which hospital police was dispatched to the unit
Fall	A patient falls, or nearly falls, in the ICU
Hemoglobin	A shift with an abrupt drop in hematocrit greater than 4 within 24 hours, given it happens after 2 hours of admission to the ward
Handoff	Errors or near misses in which information transfer between healthcare providers contributes to mistreatment of patients.
Identification	Errors or near misses in which the wrong patient was given an intervention
Lab	Errors or near misses related to lost, misunderstood, or preprocessed lab procedures
Medication	Errors or near misses related to wrong dose, or wrong medication
Safety	Errors or near misses related to safety disturbances in the unit

Table 3.1: Harms which are attributable to shifts

3.2 Development of Drivers

In this section we describe the set of drivers that are included in our model. As already denoted, the development of the set of hypothesized drivers was done in collaboration with a multidisciplinary team. The team used several approaches to hypothesize what conditions (drivers) could affect harm. First, we relied on the understanding of human errors [Reason, 1990]. Second, we reviewed a sample of over thirty past incidents to develop categories of potential drivers. Third, we used the collective clinical and operational experience of the team supported by data analysis to propose additional drivers.

The goal of this approach was to establish a set of measures that are recorded in the Electronic Medical Records at BIDMC and give insights into the nature of the work related activities, healthcare providers, and patients who are in the ICU during the period of study. We refer to these measurements as "Drivers." These drivers are categorized in four groups: Acuity, Unfamiliarity, Workload, and Other.

Measures of acuity collectively describe severity of illnesses patients. Senior clinicians from BIDMC hypothesized that high levels of acuity may be correlated with adverse outcomes. It may be the case that high levels of acuity may expose healthcare practitioners to greater rates of knowledge-based errors because they are caring for complicated patients with multiple complications with no clear standard of care.

Drivers related to unfamiliarity were developed to capture situations which the ICU operated in irregular conditions. These conditions may lead to rules-based errors, particularly if the staff does not identify that they are performing irregular operations. These drivers may include staffing decisions that place healthcare providers in wards to which they are not usually assigned, or rare patient needs that are in some way unique or unusual.

Workload drivers measure how busy the nursing staff is during the shift. It is hypothesized in this thesis that high workloads lead to increased rates of preventable harm because rule and knowledge based errors may become more frequent as the brain becomes more heavily loaded.

3.2.1 Time Unit

As already noted, drivers and harms in our model relate to a twelve hour work period between 7am and 7pm, known as a shift. BIDMC is open for a night and day shift every day of the year. Many activities that occur in the ICU can be attributed to a person, who is schedule to work on a shift, which allows activity levels to be determined to a twelve hour granularity. Personnel change at the end of a shift, and the presence of a new team affects many of the drivers, marking a natural shift into a distinctly new environment.

There are many other activities which have recorded timestamps, such as administration of medications, but anecdotal evidence indicates that the time stamps on many records are in general inaccurate by several hours because practitioners often complete several tasks before doing their record keeping. Due to the breadth of the investigation, this inaccuracy would make automated investigation of ICU activities intractable for an analysis on an hour-by-hour basis. The investigation also considers the differences between night shift and day shift activities, implying that aggregations over multiple shifts may average away important sources of variance between shifts.

3.2.2 Acuity Drivers

In a medical setting, the word "Acuity" is often used contextually to convey several different meanings. We define "Acuity" as the severity of an illness. We use several commonly used measures of acuity in this work:

1. Sequential Organ Failure Assessment score, or SOFA. The SOFA score was developed to measure the acuity of patients in the ICU [Vincent et al., 1996]. SOFA scores are evaluated several times over a shift for patients who require mechanical breathing support called ventilators. In this work we use the maximum SOFA score on a shift, which has been shown to be a good predictor for acuity [Janssens et al., 2000]. We aggregate the SOFA score for each patient in a ward over a shift as follows:

$$\text{Ward Sofa} = \frac{1}{V}\sum_{k=1}^{V} max(SOFA_k), \tag{3.1}$$

where V is the number patients on a ventilator during the shift in the ward and $SOFA_k$ is all of the SOFA scores for patient k during the shift.

2. Length of Stay in the ICU

   The Length of Stay measures the average duration of time the patients have been in the ward since they were admitted to the ICU.

$$LS = \frac{1}{N}\sum_{k=1}^{N} MIN(\text{ES, DT})_k - \text{WAT}_k, \tag{3.2}$$

   where N is the number of patients in the ward during the shift for which the Length of Stay is calculated, ES is the time the shift ended, DT is the time and day on which the patient was discharged from the ward, and WAT (Ward Admission Time) is the time the patient was admitted to the ward. The clinical intuition from hospital staff at BIDMC suggests that both short stays and long stays could be indicators of elevated risk; short stays indicated that newly arrived patients require urgent critical care, and long stays indicated that the patient was remains too unstable to be transferred to a step-down unit. In general, long lengths of stay are considered undesirable and have been suggested as a patient level quality metric for the ICU [De Vos et al., 2007]

3. Length of Hospital Stay

   The Length of Hospital Stay measures the average amount of time the patients have been in the ward since they were admitted to the hospital.

$$LS = \frac{1}{N}\sum_{k=1}^{N} MIN(\text{ES, DT})_k - \text{HAT}_k \tag{3.3}$$

   where ES, DT and N are the same as before and HAT is the time and day the patient was admitted to the Hospital.

4. First 24 hours:

   The fraction of patients in the ICU who were admitted on shift or in the previous two shifts.

Experts in the ICU consider recent admission the unit an indicator of acuity because the ICU is designated as an acute treatment facility. Patients are sent there because they need advanced care. We compute this driver in the following way:

$$F24 = \frac{1}{N}\sum_{k=1}^{N}\delta_k, \left\{ \begin{array}{ll} \delta_k = 1 & \text{if } LS_k \leq 36 \text{ Hours} \\ \delta_k = 0 & \text{if } LS_k > 36 \text{ Hours} \end{array} \right\} \tag{3.4}$$

3.2.3 Unfamiliarity

These drivers capture shifts in the ICU in which staff dealt with unusual situations. High levels of irregular environments, patients, or procedures can place high cognitive loads on staff and allow human errors to creep into healthcare delivery.

1. Float Nurse:

   A float nurse is a nurse who is working in a ward to which they are not usually assigned. For example, a nurse hired and trained in the Surgical ICU may on occasion work in the Medical ICU. BIDMC has a "Float Nurse Pool", which consists of nurses who are trained to work in any ICU – this metric does not include nurse from this pool. The measure is the fraction of nurses in the room who are "Floating:"

$$FN = \frac{1}{S}\sum_{k=1}^{S}\delta_k, \left\{ \begin{array}{ll} \delta_k = 1 & \text{if Nurse k is ``Floating''} \\ \delta_k = 0 & \text{if Nurse k is at ``Home''} \end{array} \right\} \tag{3.5}$$

   Where S is the total number of nurses that are assigned to a shift. Float nurses are easily identified from the staffing records at BIDMC. Each record shows the shift assignments for each nurse and also the ward each nurse was actually qualified to work in.

2. New Nurse:

   Fraction of Nurses in the room who have been hired to work in the ICU's within the last year

$$NN = \frac{1}{S}\sum_{k=1}^{S}\delta_k, \left\{ \begin{array}{ll} \delta_k = 1 & \text{if Nurse k was hired within the prior year} \\ \delta_k = 0 & \text{if Nurse k was not hired within the prior year} \end{array} \right\} \tag{3.6}$$

Where S is the total number of nurses that are assigned to a shift. New nurses are identified from the staffing records at BIDMC by calculating the difference between the nurse's hire date and the date of the shift which was worked by the nurse. If the elapsed time is less than one year, the nurse is considered "new."

3. Rare Procedures:

   A rare procedure is a medical intervention that has been performed less than 90 times in a ward over the two year period. Procedures which are rarely performed require the brain to operate outside of their regular routine, exposing clinicians to rules-based or knowledge-based errors.

$$RP = \delta, \left\{ \begin{array}{ll} \delta = 1 & \text{if a rare procedure occured} \\ \delta = 0 & \text{if a rare procedure did not occur} \end{array} \right\} \tag{3.7}$$

4. Boarding Patient:

   A patient is considered "Boarding" when they are assigned to ward that does not usually care for patients with the needs described by the medical service that is providing care. Table 3.2 was developed with the aid of senior physicians at BIDMC and describes which medical services are considered at home in each ward. Electronic medical records indicate each patient's service and location, and if a patient is not in their "home" ward, they are considered "boarding"

$$BP = \frac{1}{N}\sum_{k=1}^{N} \delta_k, \left\{ \begin{array}{ll} \delta_k = 1 & \text{if patient is Boarding} \\ \delta_k = 0 & \text{if patient is not Boarding} \end{array} \right\} \tag{3.8}$$

   Definitions of "Boarding" may vary between hospitals. For this thesis, the definitions of home patients can be found in Table 3.2.

Service	Home Wards
Med	FICU, MICU
CSURG	CVICU
VSURG	CVICU
CMED	CCU
NMED	SICU, TSICU
OMED	FICU, SICU, TSICU
GU	FICU, SICU, TSICU
TSURG	SICU, TSICU
NSURG	SICU, TSICU
OBS	FICU
ORTH	FICU, MICU, SICU, TSICU
TRAUM	TSICU
GYN	FICU, MICU
ENT	SICU, MICU
PSURG	SICU, TSICU

Table 3.2: Regular ICU wards for medical services

3.2.4 Workload

It has also been suggested that nursing workload has a strong impact on safety in the ICU [Carayon and Gürses, 2005]. The following metrics were suggested by the expert staff at BIDMC as events which imply how much work is being done by the staff. In conjunction with the more rigorous nursing workload metrics these are other measures of workload suggested by the staff at BIDMC. The following measures were considered:

Nursing Workload: A great deal of effort has been extended to the measurement of nursing workload in the ICU. An exhaustive study of the time consumed by executing 76 common ICU nursing tasks was conducted in 1983 and validated in hospitals around the world [Keene and Cullen, 1983]. The scoring system was later refined to the TISS-28 [Miranda et al., 1996], which categorized and simplified the original score without significant loss of accuracy. For many ICUs the TISS-28 remained too complex to implement, and the Nine Equivalents of Nursing Manpower (NEMS) was developed in 1997 [Miranda et al., 1997] and validated in 1999 [Rothen et al., 1999]. NEMS and subsequent methods use broader and mode categorical descriptions of the ICU to measure and predict workload.

The workload score used in this work is based on the Simplified Therapeutic Intervention Scoring System, which was developed in 1996 Miranda et al. [1996]. The TISS-28 was developed to measure the nursing workload in an ICU during a shift and has been found to be useful for predicting workload, resource utilization and cost but not acuity [Hariharan et al., 2007]. A nurse can of do 46 TISS points per eight hour shift, or 10.5 minutes per TISS point. The BIDMC shifts are 12 hours, and workloads have changed since the publication of the TISS-28. Expert opinions from physician and nursing staff at BIDMC were solicited to modernize the TISS-28.

The work of Ma, 2015, was instrumental in defining the revised measures of nursing workload in the ICU. Ma and his team shadowed nurse to observed differences in patient types, nursing workflow, and staffing patterns. Through consultation from expert senior medical staff, they produced table 3.3 to measure nursing workload.

To score a shift, one must add the patient activities that occurred during the shift for all of the patient on the ward and then normalize by the total number of patients. Appendix A shows the distribution of the unit level TISS score: a shift with score of 0-18 is light workload, 18-24 is moderate workload, and 24+ is high workload.

Addition Workload Measures: The following measures were developed with the clinical insight of the professional ICU staff at BIDMC. They represent approximate measures of nursing workload in a ICU on a shift.

1. Hours of Care:

   the Hours of Care is the ratio of the total number of hours worked by nurses to the total number of hours patients spend in a bed in the ICU. A low Hours of car ratio means that nurses are taking care of multiple patients, and a high ratio means that nurses are taking care of fewer patients. Typically, BIDMC staff one nurses per one or two patients based on their clinical expectation for how much effort the patient will require.

2. Admissions:

   The fraction of total patients in a ward on a given shift who were admitted on that shift. Admissions require extra communication as the patient's condition and medical history is handed from one team of providers to another.

Basic Activities	**Points**		**Renal Support**	**Points**
Standard Monitoring (All Patients)	5		CRRT	8
Routine Lab Draw (All Patients)	1		Measuring Urine Output	2
Routine Medication (Alll Patients)	2		Diuresing (Lasix)	3
IV insulin/ meds with extensive monitoring				
Routine dressing changes (All Patients)	1		**Neurologic Support**	
Care of drains (All Patients)	3		ICP Drain	4
Pressure ulcer	1			
			Metabolic support	
Ventilatory Support			Acidosis/Alkalosis	4
On a Ventillator	5		TPN/OPN	2
O2 delivery assisstance	1		Tube Feeds	3
Has a trache	1			
Chest CT (All Patients)	1		**Specific Interventions**	
			Single Procedure done in ICU	3
Cardiovascular Support			Multiple procedures done in ICU	5
Single vasoactive medication	3		Travel (OR, Cath lab, ERCP)	5
Multiple vasoactive medications	4			
1.5L IVF/blood products per shift	4			
Arterial catheter (in access line/invasive)	2			
PA Catheter, LVAD, Tandem heart	8			
Impella,PiCO, ECMO, Alsius, Arctic Sun	8			
Heart Mate,Blakemore, Massive Transfusion	8			
Central venous line	2			
Code blue in last 24hrs	3			

Table 3.3: Scoring System for Nursing Workload

3. Discharges:

   The fraction of total patients in a ward on a given shift who were discharged on that shift. Discharges require extra communication as the patient's condition and medical history is handed from one team of providers to another.

3.2.5 Other Events

Literature reviews and clinical staff at BIDMC suggested the exploration of other indicators which don't fall into a convenient category. They are presented below:

1. Readmission:

   The fraction of patients on a shift in a ward who were discharged from an ICU within the previous 72 hours. The measure of readmission as an indicator of ICU quality was first developed by Chen et al. [1998].

2. "ED Critical":

   The fraction of patients on a shift in a ward who have the name "ED Critical." These patients have an unknown name and therefore no known medical history at the time of treatment. Medical history can be very important when determining how to care for a patient, and hospital staff may change their routine when a patient with no history requires treatment.

3. Night and Day:

   Night shifts occur between the hours of 7pm and 7am. Day shifts run from 7am to 7pm. A binary variable is included to indicate night and day shifts. Senior clinicians at BIDMC suggest that procedures and staffing patterns may be different between day and night shifts.

4. Weekend:

   Weekend shifts occur between 7pm Friday and 7am on Monday. A binary variable is included to indicate weekend and weekday shifts. BIDMC staff indicated that operational patterns may change between the regular week and the weekends.

5. Unit indicator:

   When wards are aggregated together, a categorical variable is included to indicate which ward each shift belongs to. This variable is important because it may lend insight to significant differences between ICU wards within the hospital.

Chapter 4

Statistical Methods

The challenges to developing a model to identify risky states are numerous:

1. The harm events are uncommon.

   After aggregating the harm considered in this analysis, only 3.5% of shifts have a harmful event. A statistical method which remains effective for unbalanced datasets must be considered.

2. The Risky States must be descriptive.

   There are many statistical tools, such as random forests and neural nets, that can identify clusters of high-risk shifts. While, these methods are helpful for predicting outcomes, they are not effective for understanding the underlying conditions that lead to harm, and do not facilitate the development of risk mitigation strategies. A tool must be used which facilitates understanding the conditions that lead to the development of high levels of risk.

3. The problem is high dimensional.

   With over 20 independent variables and nearly 8700 observations, not all analytic methods are appropriate; as the number of independent variables grows, the dimensionality and thus volume of space increases rapidly. Within this large space, even 8700 data points can quickly become a sparse number of observations, and many tools, such as multiple regression, will lose their statistical power. Fortunately, there are several machine learning algorithms designed to work within this challenge.

4. The results must be statistically robust.

   To our knowledge, this is the first attempt to identify risk in ICU wards at an aggregated level. A new model which predicts risk should be consistent with other methods and withstand cross-validation from multiple techniques.

Classification Trees are a useful supervised learning algorithm that are efficient for working with large numbers of independent variables. Classification trees are a good approach because they consider only the most important driver interactions and help the user identify the groups of drivers that best estimate risk. Unfortunately, for some datasets classification trees become unstable so great care must be taken to validate the results of this highly adaptive learning algorithm. K-Nearest Neighbour (K-NN) methods are also helpful for defining probability density functions in high dimensional space and are used in this thesis to validate the Classification Tree method.

The method used in this work is a combined use of Classification Trees and the K-Nearest Neighbour. A brief outline of this chapter and statement of the algorithm is below:

1. Run multiple classification trees to reduce problem dimensionality.
2. Calibrate and create "Risky States" with a Classification Tree to identify high and low risk states.
3. Determine the stability of the tree by comparing the results with a second machine learning algorithm, the K-Nearest Neighbour.
4. Determine the statistical significance of the Risky States with the Man-Whitney-Wilcoxn hypothesis test.

4.1 Classification Trees

In this thesis, Classification trees are used as a clustering recognition algorithm to identify risky states. Classification trees can be thought of as a method for searching a hypothesis space for clusters that best fit the training data. A good fit is a tree that best separates observations of harmful events from observations without harmful events. The hypothesis space is the complete space of finite values for the drivers, and comprises every observed state of the ICUs. The algorithm iteratively searches the hypothesis space for driver values (thresholds) which best split observations of regions in hypothesis space where harm occurred from those where it did not. The classification tree will continue to split the hypothesis space into smaller regions ("Risky States") until the maximum number of splits is reached, the minimum number of observations which define a region remains, or the region is of a single type of observed outcome (harm or no harm). Breiman et al. [1984] provides the full details of how the classification tree algorithm works.

4.1.1 Overview

Classification trees were originally developed as method to the predict the likelihood that a new observation will belong to a class. A classification tree belongs to the subset of machine learning algorithms called 'supervised learning' because the training observations are first marked by their class before the algorithm is run. Classification Trees differ from Regression trees in that Regression Trees are trained on continuous targets. Classification trees have several important features that make them advantageous for this investigation:

1. Non parametric

   This method makes no assumptions about the shape of the driver data. Thus the method works well with many types of data distributions that could be underlying in the various drivers such as long tailed distributions , multi-modal distributions, and outliers.

2. Robust treatment of categorical variables

   Many drivers under investigation are categorical, such as nigh or day, or ICU type. Classification tree methods work well with categorical variables.

3. Discovers "interactions" among variables

   The classification tree estimates a probability of an event occurring based on the values of **all** the drivers that define the tree.

4.1.2 Criticisms

A significant drawback to classification trees is that they can be sensitive to perturbations in the data that they are trained on. This means that a small differences in harm distributions has the potential to create differently shaped trees. The large dimensionality of the current problem poses the possibility that weak drivers could be falsely selected. While ensemble methods such as AdaBoost and Random Forests have been developed to mitigate these issues, these "black box" techniques do not provide clearly defined risky states around which a management strategy can be built.

4.1.3 Classification Tree for Feature Selection

To mitigate the instability issues of Classification Trees, an experiment was conducted to better understand the instability issues with the current dataset. Instantaneous harms and drivers for five wards (MICU, SICU, FIN4, TSICU and CCU) were randomly split into 10,000 training sets, with each training set containing 50% of the available data. Classification trees were built on the each of the subsequent training sets, and the total frequency with which drivers appear shown in Figure 4-1. The figure shows that some drivers are selected significantly more often than others. We conservatively propose that drivers which appear in more than half of the classification trees may have a significant impact on predicting risky states, and include these variables in the K-NN validation method.

4.1.4 Classification Tree for Risky States

In this thesis, Classification trees are used to identify clusters of shifts that have similar drivers and rates of preventable harm. Using retrospective data, if a preventable harm from Table 3.1 occurred on a shift, that shift is marked. The shifts are parametrized by their drivers, which are projected into the hypothesis space. The Classification tree algorithm is then run to determine multiple thresholds in the hypothesis space that best separates shifts on which harms occurred from those on which it did not. The thresholds are hyperplanes which divide the hypothesis space

Frequency.png

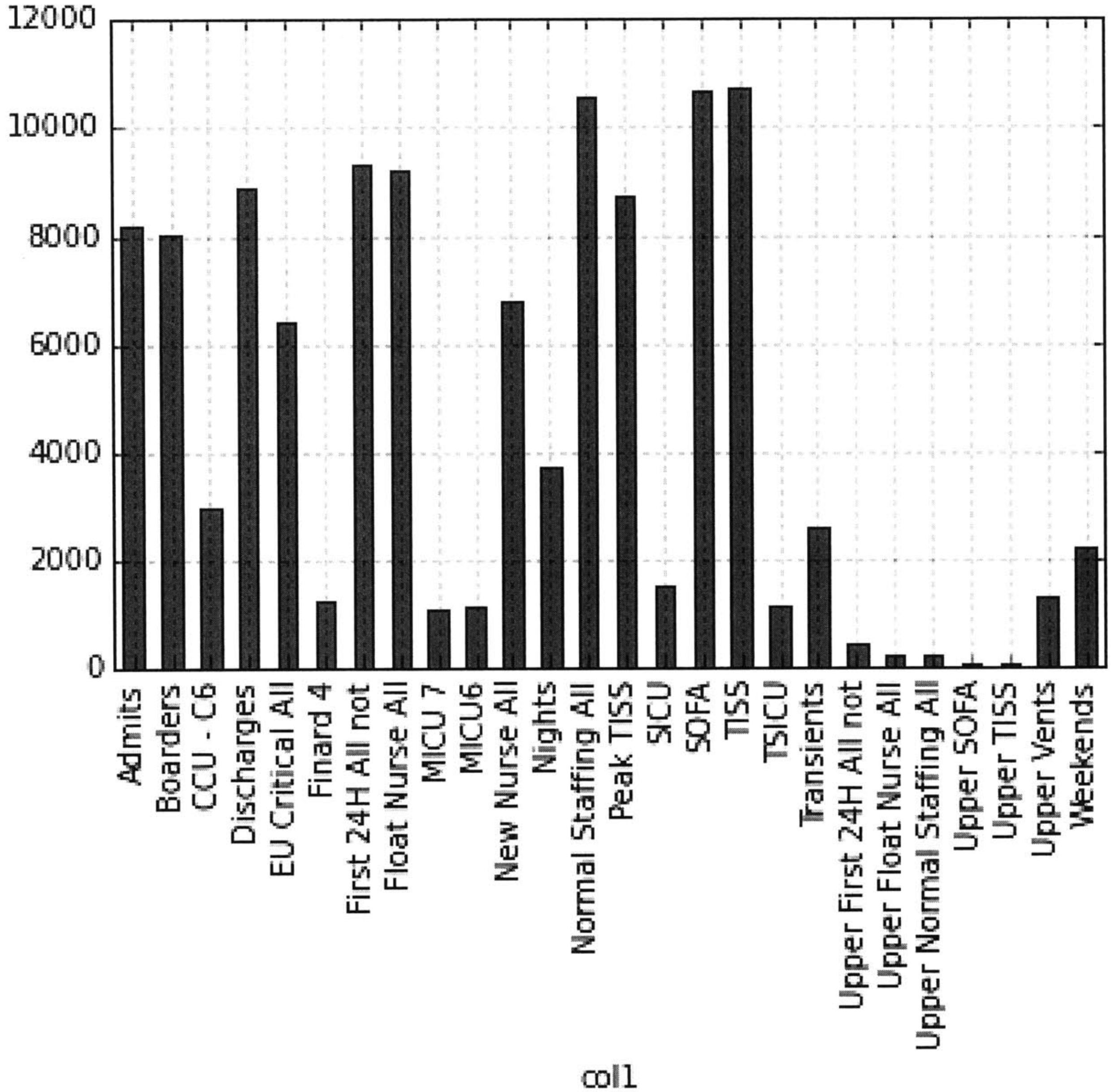
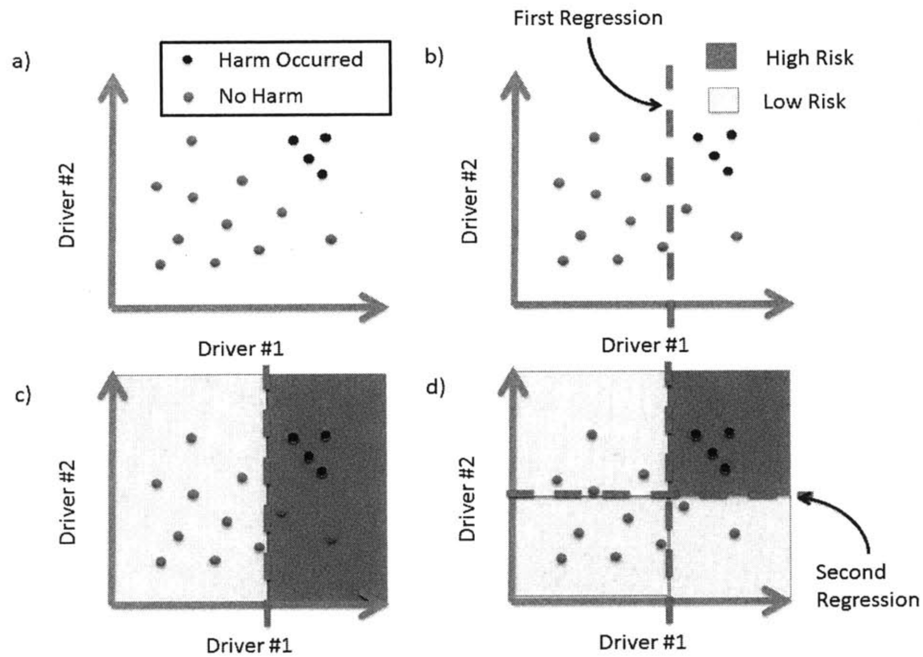


Figure 4-1: Greater frequencies indicate stronger predictive power

into partitioned volumes, and the volumes are known as "Risky States". Figure 4-2 illustrates the process in two dimensions for a generic set of drivers:

A discussion of the algorithm and details of its implimentation can be reviewed at Pedregosa et al. [2011]



Tree.png

Figure 4-2:
Step a) Classify shifts on which harm occurred.
Step b) find best plane that separates classes
Step c) A high risk region is separated from a low risk region by a threshold value for Driver #1
Step d) repeat until terminal conditions are met

4.2 K-Nearest Neighbour

4.2.1 Overview

The K-Nearest Neighbour Method (K-NN) is a non-parametric algorithm for estimating classification and regression. When K-NN is used for classification, as in this thesis, the probability that an observation belongs to a class is determined by taking the average of the classes for the k nearest training observations. A small k number will make the method sensitive to noise and a large k decreases the sensitivity of this method. It is generally suggested that $k = 30$ provides a good balance between the two.

The K-NN distances are estimated based on euclidean norms for all variables that are input into the model. Variables which are random, or unpredictable, can therefore drag unrelated observations into the nearest neighbour radius. For this reason, we train the K-NN model only on drivers which have been selected as important features for the driver frequency experiment in the previous section.

4.2.2 Use of K-NN

The K-NN model is run first on the set of drivers (and associated harms) that were identified in the results of Section 4.1.3. This is used to create a probability density function, ρ, for the entire population of shifts.

Next, the K-NN method is run on each of the risky states that were proposed by the classification tree. This provides an alternative method for estimating the probability of harm in each of these states μ_β as well as a probability density function for each state, ρ_β. These metrics will be useful for cross-validated the Risky States.

4.3 Model Metrics: Separation, Instability, Stability and Significance

It is generally accepted that a "good" mathematical model will tell the investigator something that they didn't understand before beginning an investigation. In this setting, it is trivial to estimate the a priori probability of harm occurring on a shift, μ, from the simple formula

$$\mu = \frac{\sum_{n=1}^{N} \delta_n}{N}, \left\{ \begin{array}{ll} \delta_H = 1 & \text{if a harm occured on shift n} \\ \delta = 0 & \text{if a harm did not occur on shift n} \end{array} \right\} \tag{4.1}$$

where N is the total number of shifts.

In this work, $\mu = 3.5\%$. This thesis defines information gain as the number of shifts which are different from our a-priori knowledge about them. In developing this model, several important measures for the usefulness and reliability of the output are defined.

4.3.1 Separation

Separation (σ) is a metric to describe information gain. Information gain is a measure of the number of shifts which can be separated from the population average in a meaningful way. Mathematically,

$$\sigma = \begin{cases} \sum_\beta \frac{n_\beta}{N} \frac{\mu - \mu_\beta}{\mu} & \text{if } \mu > \mu_\beta \\ \sum_\beta \frac{n_\beta}{N} \frac{\mu_\beta - \mu}{1-\mu} & \text{otherwise} \end{cases} \tag{4.2}$$

where $\mu_{\beta,\kappa}$ is the probability of harm predicted for state β by K-NN method and $\mu_{\beta,\tau}$ is the probability of harm predicted by the Classification Tree method, n_β is the number of shifts in state β and μ is defined in equation 4.1.

This metric has the desirable properties of becoming larger as more observations are separated from μ. By construction, this metric ranges between zero and one. It is adjusted for bias in μ so that information gain for datasets with $\mu \neq .5$ will not be overweighted.

The measurement is non-singular, which means that several models with different shapes can have the same score.

In the figure below, the separation metric is calculated for classification trees to demonstrate the effect of varying the maximum depth of the tree ($0 - 25$) and the minimum number of shifts that are required to characterize a risky state (30-300).

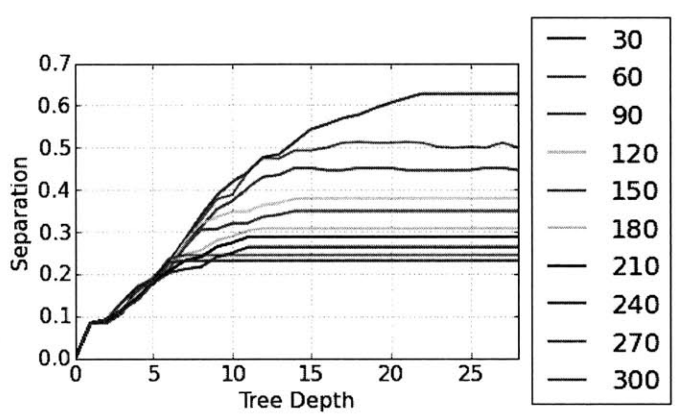


Figure 4-3: Classification Tree Parameter Search

The maximum depth of the tree and the minimum number of shifts per state are the two major features in tuning a classification tree. By varying these tuning parameters, we are able to gain the following insights:

1. Changing tree depth reveals that separation, in general, increases with the depth of the tree. As the tree continues to grow in depth the separation levels off, at which point all of the shifts have been assigned to a state.

2. The inverse relationship between minimum state size and separation can be seen by looking at the vertical distance between trend-lines at any particular tree depth. As the minimum shifts per state shrinks, separation increases.

3. The metric is non-linear

This metric alone tells us that separation increases with smaller minimum state sizes, and with greater tree depths. Taken to the limit, a model with states of 1 and unlimited depth would be useless in its ability to predict the future and extremely unstable for predicting the risk of nearby states. Where must therefore temper the desire for separation with a robust treatment of instability.

4.3.2 Instability

In this thesis, instability is defined by comparing different predictions for harm rates that are generated by the two different machine learning methods.

$$\Delta = \frac{1}{N} \sum_{\beta} n_{\beta} |\mu_{\beta,\tau} - \mu_{\beta,\kappa}| \tag{4.3}$$

For each state that is proposed by the classification tree method, the difference in the harm rates that are predicted by the K-NN method and the Classification method are accumulated. A larger Δ indicates that the two methods have very different predictions about the level of harm in many of the states. A smaller Δ indicates greater agreement.

As before, we seek to understand the influence of the classification tree tuning parameters, Tree Depth and the minimum number of shifts that define a state and conduct a parameter search:

The figure above demonstrates four important behaviours:

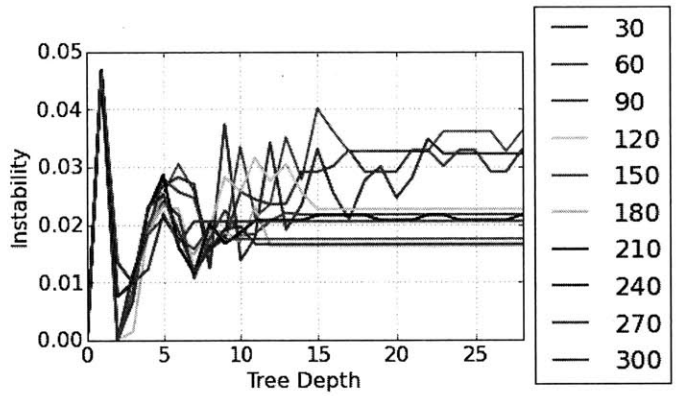


Figure 4-4: Classification Tree Stability Search

1. Oscillations in instability dampen as tree depth increases. This suggests that as the tree grows, unstable states are generated and then split into stable states with fewer samples per state.

2. Instability is inversely proportional to minimum number of shifts per state. When a smaller number of samples are allowed to define a state, the instability of the model grows. This is in direct opposition to the stability metric.

3. Instability increases with tree depth. Although the oscillations dampen with depth, the bias increases.

4. The metric is non-linear

The apparent opposition between Instability and Separation yields rapid analytic insight into the trade-off between tree depth and minimum samples per state. Trees of depth eight seem to converge, suggesting good stability at that level between under-fitting and over-fitting. At tree of depth eight, minimum state size of 120 appears large enough not to suffer from the instability and small enough to provide good separation.

4.3.3 Second Stability Check

The second stability check is an additional filter to check the quality of a state determined by the Classification Tree. A state is unstable if one model classifies the mean of a state as elevated risk, while the other model classifies as low risk.

$$f(\mu, \mu_{\beta,\kappa}, \mu_{\beta,\tau}) = \begin{cases} \mu_{\beta,\kappa} < \mu \text{ and } \mu_{\beta,\tau} > \mu & = 0 \\ \mu_{\beta,\kappa} > \mu \text{ and } \mu_{\beta,\tau} < \mu & = 0 \\ \text{else} & = 1 \end{cases}$$

States become contradictory when the thresholds that define a state are poorly constructed. The probability density functions that are defined by the K-NN method are trained on the the nearest neighbours. If the neighbours are across the threshold defined by the Classification Tree, they are included in the K-NN model when the mean of the bucket is calculated. Thus, states with different classifications near the threshold will affect the K-NN model, but not the classification tree. This filter is effective because it eliminates two types of false states:

1. A state near the mean, but on different sides of it according to the two models, is not an important state because it yields no new significant information beyond the a priori assumptions.

2. A state which is far from the mean and dragged across the mean by its neighbours is "boxed in" by nearby members of a different class and is likely over-fit.

4.3.4 Significance

The final check to understand the stability of the risky states that are proposed by the Classification Tree is a test for statistical significance. It is important that the distribution of harm in the risky state is different than the distribution of harm in the aggregated observations of shifts. If the two distributions are not different, then the algorithm has not provided new information about harm in the state.

The Mann-Whitney-Wilcoxon (MWW) test is a non-parametric statistical test to determine the significance of the null hypothesis that two samples come from the same population versus the alternative hypothesis that the two samples are from different populations. The distributions of harm for each state and for the general aggregation of all states are determined with the K-NN method.

In this thesis, the alternative hypothesis, H_1, is that the mean of harm, $\mu_{\beta,\kappa}$, in a particular state, β, is different than the mean of harm in the general population, μ. We are interested in cases for which $\mu_{\beta,\kappa} > \mu$ and $\mu_{\beta,\kappa} < \mu$. The null hypothesis is $\mu_{\beta,\kappa} = \mu$. For this test, p-values of less than .05 indicate significant distributions.

Chapter 5

Results

5.1 Summary of Results

A summary of sixteen states which are of interest can bee seen in Table 5.1. Each of these states passes all three tests for stability in chapter 3. A full review of these states follows in this chapter, which supports three major conclusions:

1. The root indicator for most elevated risk states is the Nursing Intensity Score. Levels of nursing workload that are even slightly above the mean are a precursor to many risky states. However, it should be noted that not all of the high risk states in Table 5.1 require combinations of drivers. No single driver is sufficient to define a high risk state.

2. Groups of newly hired nurses, in combination with elevated workload, are strong indicators of elevated risk.

3. When multiple risks factors are high, so are harm rates.

4. Shifts with low average workloads are relatively safe, and remain safc when perturbed by a single environmental factors that stress the system.

In the next sections, the definition of each individual state is presented. The proportion of each type of harm in each state is compared to the proportion of harms in the general population of

State	μ_T	Shifts	μ_K	P Value	Risk Level
12	0.019	253	0.028	4.099e-04	Low
19	0.026	1957	0.033	2.064e-02	Low
23	0.091	142	0.042	8.330e-03	High
26	0.077	206	0.053	4.553e-12	High
32	0.007	134	0.017	5.088e-11	Low
39	0.005	179	0.024	2.876e-06	Low
42	0.016	120	0.025	1.502e-03	Low
43	0.000	261	0.017	7.161e-37	Low
46	0.014	202	0.030	7.855e-03	Low
47	0.004	224	0.029	9.808e-03	Low
54	0.075	146	0.060	4.525e-13	High
58	0.040	174	0.047	9.968e-06	High
59	0.118	144	0.056	1.050e-09	High
60	0.111	153	0.058	8.487e-13	High
65	0.062	129	0.045	6.316e-04	High
73	0.072	124	0.050	1.535e-05	High

Table 5.1: Summary of Significant Outputs

harms for all states to see if certain states have a higher prevalence of particular risks. It is shown that medical errors occur disproportionately when multiple drivers of risk are elevated and hand-off errors occur disproportionately when workloads are modestly elevated and combined with a second elevated driver.

5.2 Risky States: High Risk

In the following section, eight high risk states, representing 13.2% of the shifts, are presented. The first set of states are grouped together because they describe situations that could lead to errors related to choosing the wrong rule to act on in a given situation due to moderate workload, unfamiliarity with the situation and high cognitive demand. The second set of risky states are all high risk states, and have a disproportionally higher rate of hand-off errors.

5.2.1 Rules related Errors

We begin our understanding of the features of high risk states with state 23, which from table 5.1 includes 123 shifts with a 9.1% chance of harm, more than double the apriori (3.5%).

The bar graph in the figure shows the proportion of harm in the entire set of shifts compared to the proportion of harm in state 23. The distributions are normalized by the total number of harms so that an accurate comparison for the frequency of each harm type can be better understood. State 23 shows higher frequency of arrests, falls and identification errors with proportionally fewer medication and handoff type errors than exist in the general population of shifts.

23.png

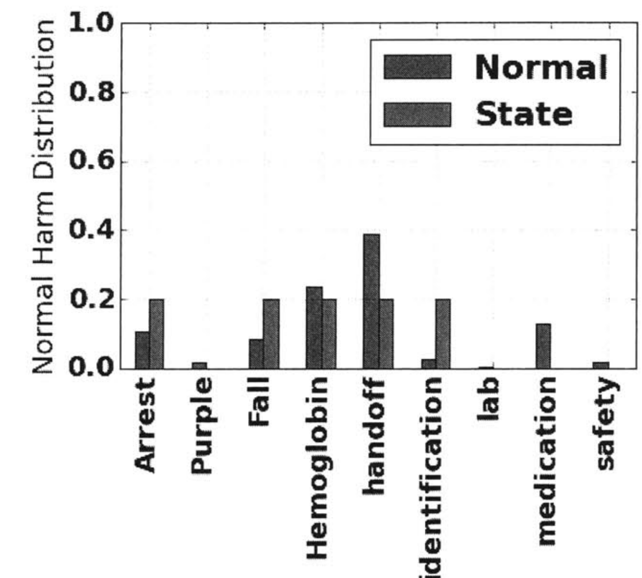


Figure 5-1: State 23 - 9.1% Chance of Harm, 142 Shifts

Driver	Bottom Threshold	Top Threshold
Nurse Intensity	21.81	24.58
SOFA	5.8	13.4
Float Nurses	72%	
Jon Doe		31.7%

Table 5.2: Midrange work with Float Staff and Jon Doe

From the table we see that state 23 suggests that on shifts with midrange workloads with midrange acuity, and over 72% of the staff working outside of their home unit. With up to 32% of the patients in the room unidentified. As one might expect, the harm breakdown in the figure shows that the number of errors related to patient identification is elevated far above the general population. The combination of moderately stress working environment with many clinical staff unfamiliar with their environment suggests that rule-relate cognitive errors are being made in greater frequency. Elevated rates of identification errors support this claim, which is that the staff is performing procedures with which they are familiar but perhaps not appropriately adjusted to their new environment an unidentified patients.

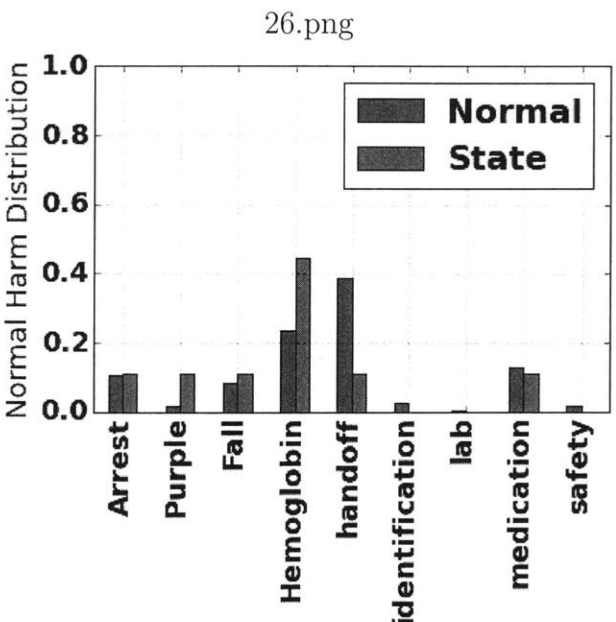


Figure 5-2: State 26 - 7.7% Chance of Harm, 206 shifts

Driver	Bottom Threshold	Top Threshold
Nurse Intensity	19.46	24.58
SOFA	5.8	13.4
Float Nurses		72%

Table 5.3: Midrange work and Acuity with Float Staff

State 26 shows a similar situation, with midrange acuity and the Nursing Intensity score also in a middle range, though slightly lower. The float nurse upper threshold is fairly high with a large range, indicating that it may not be a useful predictor for this state. The harm breakdown suggests an increase in bleeding and code purple events, but is difficult to tie these events directly to the wide range of float nursing scenarios spanning the 206 shifts. Regardless of our interpretation of the float nurses, we see that state 23 and state 26 both implicate elevated regions of risk in regions of average workload and intensity.

In state 65, the nursing workload intensity is elevated and there are large proportion of new nurses in the unity. The combination high workloads and staff that are unfamiliar with their tasks, we see elevated harm rates perhaps due to the increased frequency with which experienced nurses have to help less experienced staff, or a rise in situations where inexperienced staff are asked to perform tasks for which they are not fully prepared.

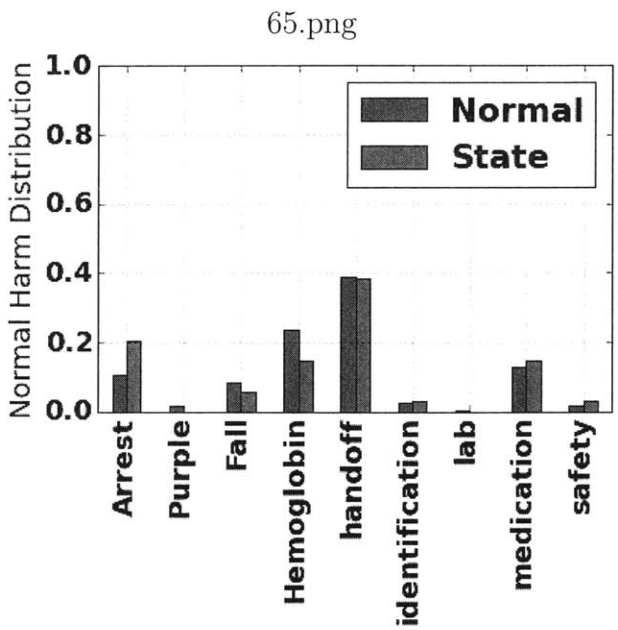


Figure 5-3: State 65 - 4.5% Chance of Harm

Driver	Bottom Threshold	Top Threshold
Nurse Intensity	24.59	
New Nurse	23.4%	
Pt : Nurse		2 : 1

Table 5.4: State 65: Busy, Multitasking and New

5.2.2 Handoff Errors

States 54, 58, 59 and 60 are similar and suggest the influence of midrange surge in the Nursing Intensity Score. The 75th percentile for nursing workload is given appendix A at 24.5. Handoffs are a rules-based task, but during these busy times, a larger number of handoff errors are made potentially due to increased cognitive loading from elevated working demands.

State 58 and State 59 are separate states but nearly identical, differentiating only in the nursing intensity score. These two states can be considered together because upper threshold of nursing intensity for state 58 is the same as the lower threshold for nursing intensity in state 59, and all other thresholds are the same. Combining these states reveales higher handoff errorsm which is why they are grouped together in this section.

5.2.3 System Breakdowns

State 73 is a set of very uncomfortable shifts in the ICU, with many out of the ordinary factors. Large numbers of new nurses, float nurses and boarders sets the stage for all three types of cognitive errors. Combining this state with workloads that are elevated results not only in a significantly increased

54.png

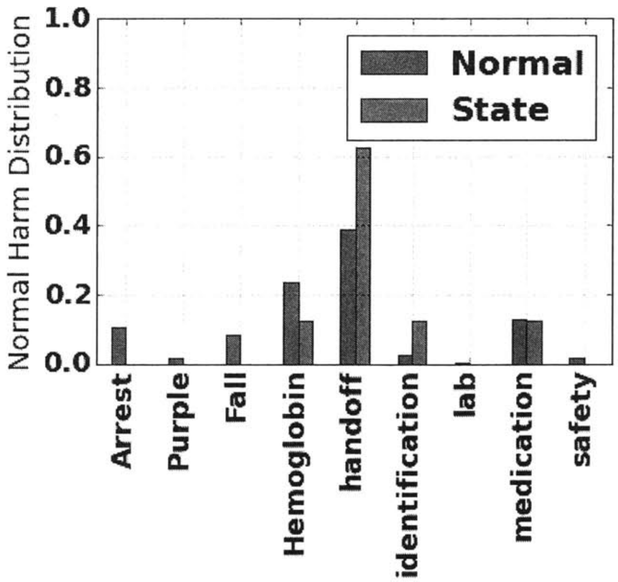


Figure 5-4: State 54 - 7.5% Chance of Harm, 146 Shifts

Driver	Bottom Threshold	Top Threshold
Nurse Intensity	24.58	29.46
New Nurse		23.4%
Discharge		26.1%
First 24H	40.8%	
SOFA		12.8
Ventilators		75%

Table 5.5: State 54: Mid Intensity Surge

58.png

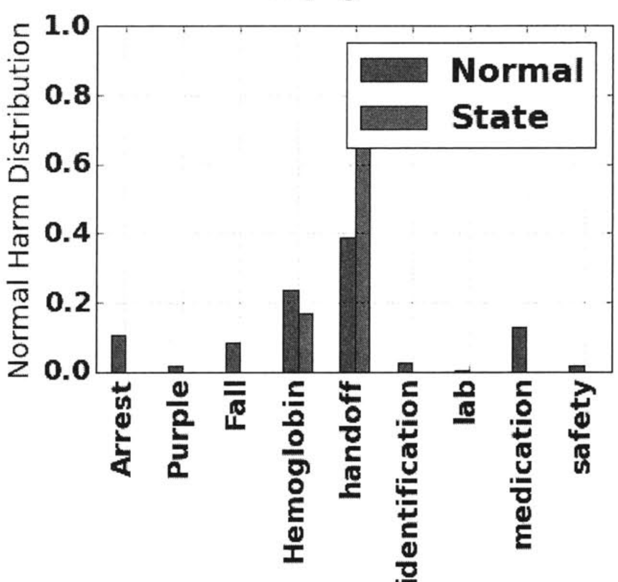


Figure 5-5: State 58- 4.0% Chance of Harm, 174 Shifts

Driver	Bottom Threshold	Top Threshold
Nurse Intensity	26.69	29.46
New Nurse		23.4%
Discharge		26.1%
First 24H		40.8%

Table 5.6: State 58: Mid Intensity Surge

likelihood of errors, but also a significant increase in patients receiving the wrong medication in particular. This state, above all other risky states, represents a large number of things gone wrong

59.png

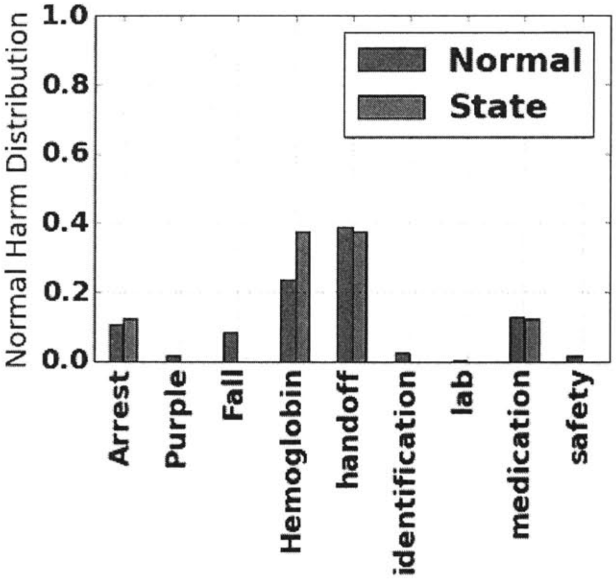


Figure 5-6: State 59- 11.8% Chance of Harm, 144 Shifts

Driver	Bottom Threshold	Top Threshold
Nurse Intensity	24.59	26.69
New Nurse		23.4%
Discharge		26.1%
First 24H		40.8%

Table 5.7: State 59: Mid Intensity Surge

60.png

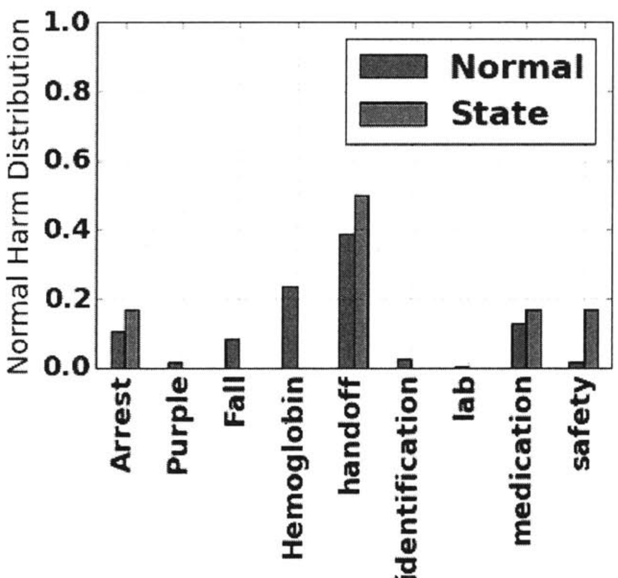


Figure 5-7: State 60- 11.1% Chance of Harm, 153 Shifts

Driver	Bottom Threshold	Top Threshold
Nurse Intensity	24.59	26.69
New Nurse		23.4%
Discharge	26.1%	

Table 5.8: State 60: Mid Intensity Surge

in the ICU.

73.png

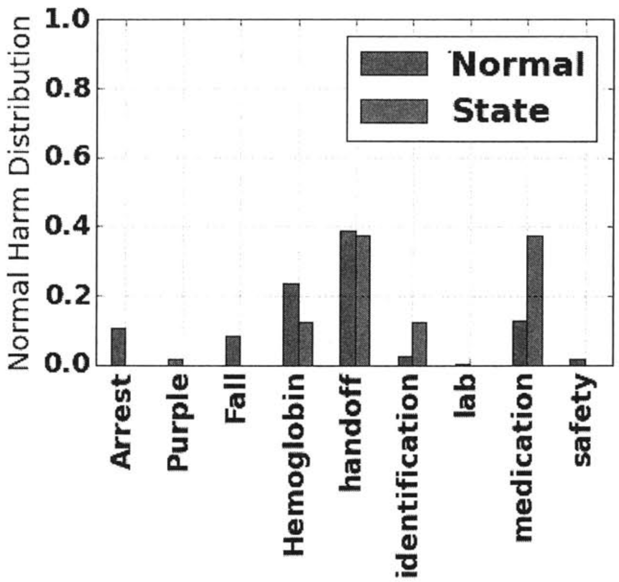


Figure 5-8: State 73- 7.2% Chance of Harm, 124 Shifts

Driver	Bottom Threshold	Top Threshold
Nurse Intensity	24.59	
New Nurse	23.4%	
Pt : Nurse	2: 1	
Float Nurse	21.1%	
Boarders	76.4%	
SOFA		11.76

Table 5.9: State 73: Busy, Multitasking and Unfamiliar

5.3 Risky States: Low Risk

There are eight states comprising 3330 shifts, or 38.2% of the operating period, with significantly less risk of harm. These shifts are characterized by medium to low average workloads, low numbers of new patients, and low patient turnover in the wards. It is interesting to note that for many of theses states, one or in some cases two risk drivers may be elevated, but the ICU delivery system remains safe in spite of single point perturbations.

State 32 has a very light workload. The apparent prevalence of handoff errors is misleading. The relative occurrence of hand-off errors is 100%, but the chance that they occur is very low at 0.7%.

State 42 and 43 are both low risk and differ only on the discharge threshold. We can see from the isolated influence of discharges switching thresholds that the risk of harm nearly doubles under this effect alone.

State 47 demonstrates a balance between low nursing intensity and high sofa scores. Acute patients alone to not make for high risk states.

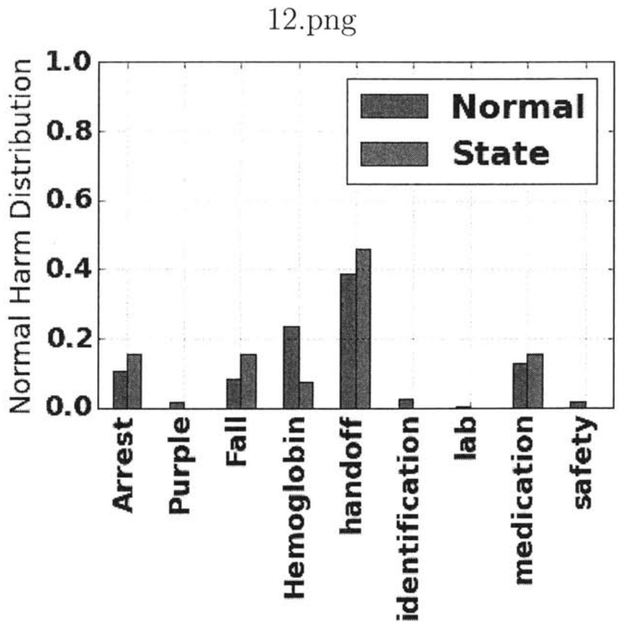


Figure 5-9: State 12 - 1.9% Chance of Harm, 253 Shifts

Driver	Bottom Threshold	Top Threshold
Nurse Intensity		24.59
SOFA	5.84	13.44
Float Nurse		5%
EU Critical		31.7%
Pt:Nurse Ratio		5:4
New Nurse	21.1%	

Table 5.10: State 12: Light Day in the ICU

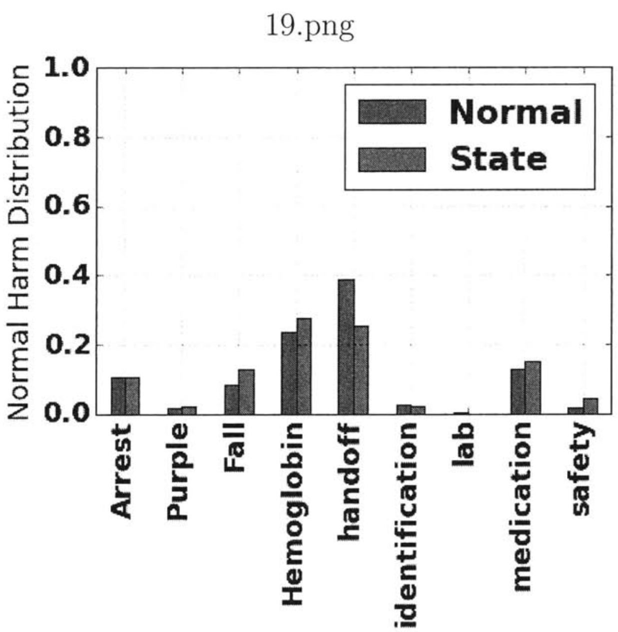


Figure 5-10: State 19 - 2.6% Chance of Harm

Driver	Bottom Threshold	Top Threshold
Nurse Intensity		24.59
SOFA	5.84	13.44
Float Nurse		72.1%
EU Critical		31.7%
Pt:Nurse Ratio	5:4	
Discharges	21.1%	29.3%

Table 5.11: State 19: 1957 shifts, Largest ICU State

32.png

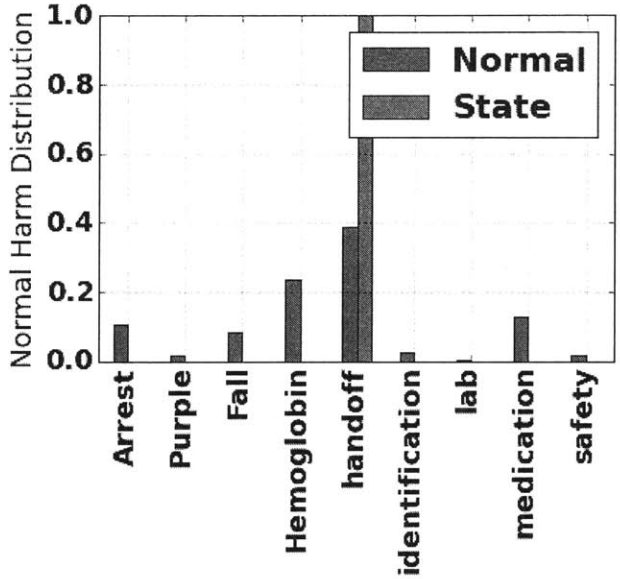


Figure 5-11: State 32 Harm Breakdown

Driver	Bottom Threshold	Top Threshold
Nurse Intensity		17.91
SOFA	13.44	14.65
Nursing Peak		23%
Pt:Nurse Ratio		5:4

Table 5.12: State 32 - 0.7% Chance of Harm, 134 Shifts

39.png

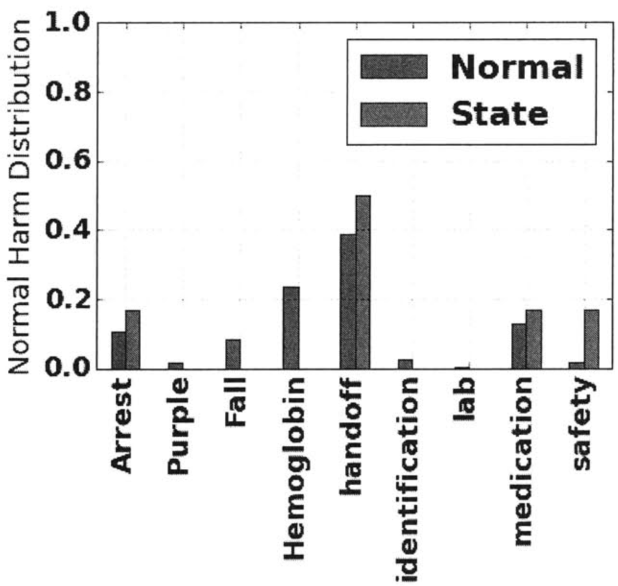


Figure 5-12: State 39 Harm Breakdown

Driver	Bottom Threshold	Top Threshold
Nurse Intensity		23.05
SOFA	13.44	
Nursing Peak	26%	
Pt:Nurse Ratio	5:4	
Boarders		89%
Discharges		10%

Table 5.13: State 39- 0.5% Chance of Harm, 179 Shifts

42.png

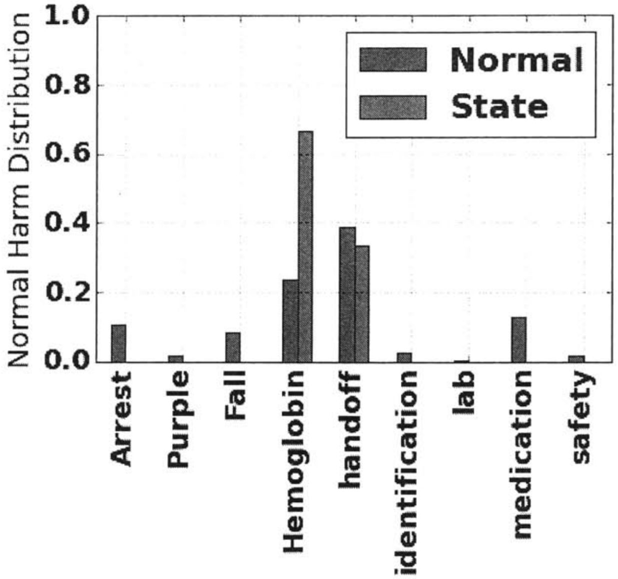


Figure 5-13: State 42 Harm Breakdown

Driver	Bottom Threshold	Top Threshold
Nurse Intensity		23.05
SOFA	13.44	
Nursing Peak		26%
Pt:Nurse Ratio		4:3
Boarders	89%	
Discharges		10%

Table 5.14: State 42- 1.6% Chance of Harm, 120 Shifts

43.png

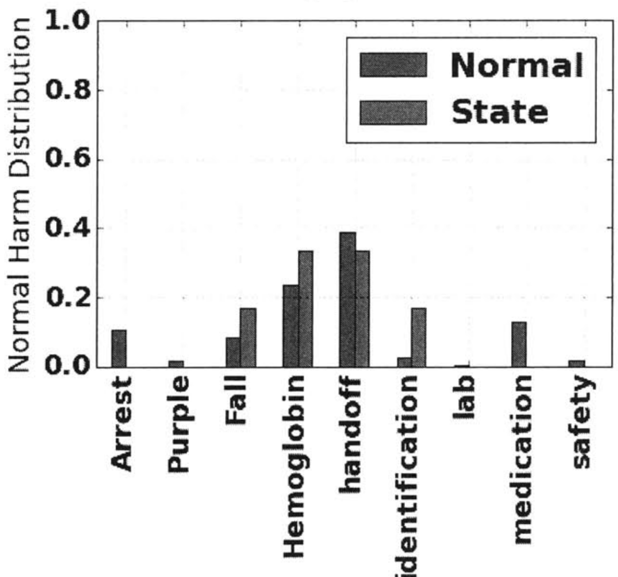


Figure 5-14: State 43 - 0.00% Chance of Harm, 261 Shifts

Driver	Bottom Threshold	Top Threshold
Nurse Intensity		23.05
SOFA	13.44	
Nursing Peak		26%
Pt:Nurse Ratio		4:3
Boarders	89%	
Discharges	10%	

Table 5.15: State 43: Quiet in the ICU

46.png

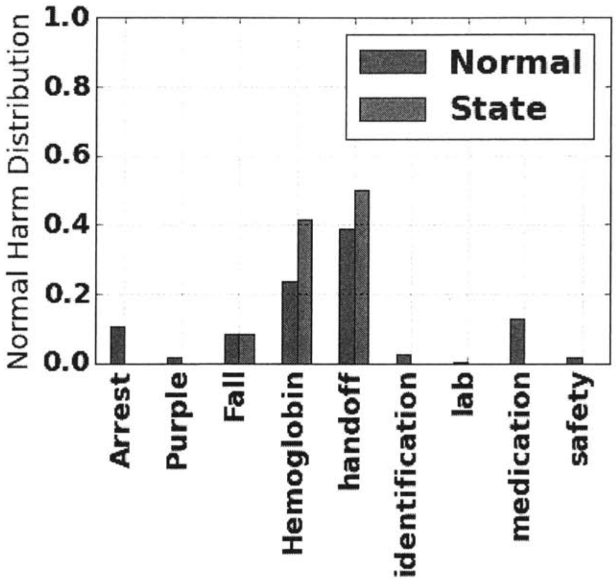


Figure 5-15: State 46 - 1.4% Chance of Harm, 202 Shifts

Driver	Bottom Threshold	Top Threshold
Nurse Intensity	20.08	23.05
SOFA	13.44	
Nursing Peak		26%

Table 5.16: State 46: Quiet in the ICU

47.png

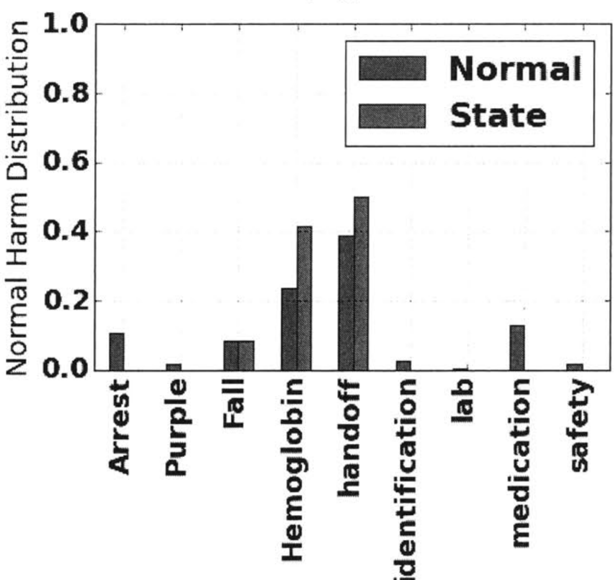


Figure 5-16: State 47 - 0.4% Chance of Harm, 224 Shifts

Driver	Bottom Threshold	Top Threshold
Nurse Intensity	23.05	24.58
SOFA	13.44	

Table 5.17: State 47: Quiet in the ICU

Chapter 6

Discussion

6.1 Conclusions

A new methodology for aggregating rare events to understand risk drivers at a systematic level has been developed. The proposed "Risky States" satisfy both statistical rigour and intuitive inspection. The "Risky States" reveals a critical care system which is well designed to operate under single sources of systemic stress but suffers from increased error rates as multiple stressors are applied.

This thesis satisfies major components of the three output of the project from Chapter 1. The measurement strategy is presented in chapter 3. Chapter 5 presents the results described in output 2 of the Moore Grant Proposal. Throughout the project, deliveries of section of code to Aptima, a software engineering company and strategic BIDMC partner, have resulted in a prototype user interface based on this work which can be installed in the BIDMC network. Further spread of these ideas had been arrange through outreach with the Veterans Affairs Hospital System, whose central quality control group has pledged the necessary resources to validate and transfer this technology to their network of hospitals.

6.2 Next Steps

The drivers presented in this thesis suggests common drivers for several types of harm. Understanding the influence of these environmental states with elevated risks sets the stage for a wide array of possible interventions. Typical risk management practices suggest "de-risking" states with a variety of process improvement techniques.

1. Mitigate. Risk mitigation can be accomplished by investing in new work-flow patterns to reduce load, more reliable treatments, or technologies that reduce workload to make healthcare delivery easier and safer.

2. Avoid. It may be possible to reduce risk levels by diversifying the patient mix in each ICU ward. Avoiding a high risk state can be accomplished with strategic staffing and patient placement. The ability to predict, within reasonable bounds, the needs of the patients and the availability of staff is a necessary capability to avoid elevated risk states.

3. Ignore. It may also benefit the risk management process to temporarily ignore the lower risk states and focus limited resources on reducing the impact or occurrence of higher risk states.

Investigating the cost and impact of these next steps, along with a pilot implementation to validate estimates, will create actionable impact on hospital operations.

6.3 Further research

To the author's knowledge, this is the first attempt to aggregate hospital errors and seek to explain them with environmental level drivers. Over the course of this research, several additional drivers may be considered important to be investigated in the future.

1. Extension to additional harms

2. Validation at other medical centres. Each medical center has potentially different environments with different capabilities to measure them. This thesis provides a set of adaptable

techniques that scales well with additional drivers and can make accurate assessments of hospitals with different characteristics.

3. Physician involvement. Due to unavailability and unreliability of electronic medical record keeping, the physician, or team of physicians, responsible for the care of patients in the ICU was in-determinant. While surprising to most readers, it is the opinion of the staff that such records are not obtainable at all similar hospitals

4. Staff Training. Detailed records on the proficiency of staff members with procedures that they are carrying out, either by training or experience, were unobtainable for this study.

5. Extension to prospective risk assessment. The current study classifies harm rates for retrospective events. The ability to predict the drivers for future states is an important capability for resource planning purposes.

Chapter 7

Hyperplane Cutting: an Efficient Algorithm for Simultaneous Selection of Machines and Optimization of Buffers During Manufacturing System Design

We depart from healthcare analytics to consider the problem of machine selection in the design of a manufacturing system. The goal is to choose machines and buffer sizes to maximize the profit generated by the line. There are more than one candidate machines for some of the process stages. The machines differ by reliability and capital cost. For long lines or cases with many stages that have many possible machines to choose among, the number of candidate machine sets can be too large to permit the problem to be solved by evaluating all possibilities. Furthermore, the evaluation of a machine set can be expensive because each machine set has a different optimal set of buffer sizes, and determining that set takes substantial computer time. We describe a heuristic algorithm which does not require an intelligent initial guess. In a set of experiments, it converged in almost all cases to the optimum, and it converged in seconds or minutes in many cases in which evaluating all possibilities would have taken years.

7.1 Introduction

7.1.1 Problem

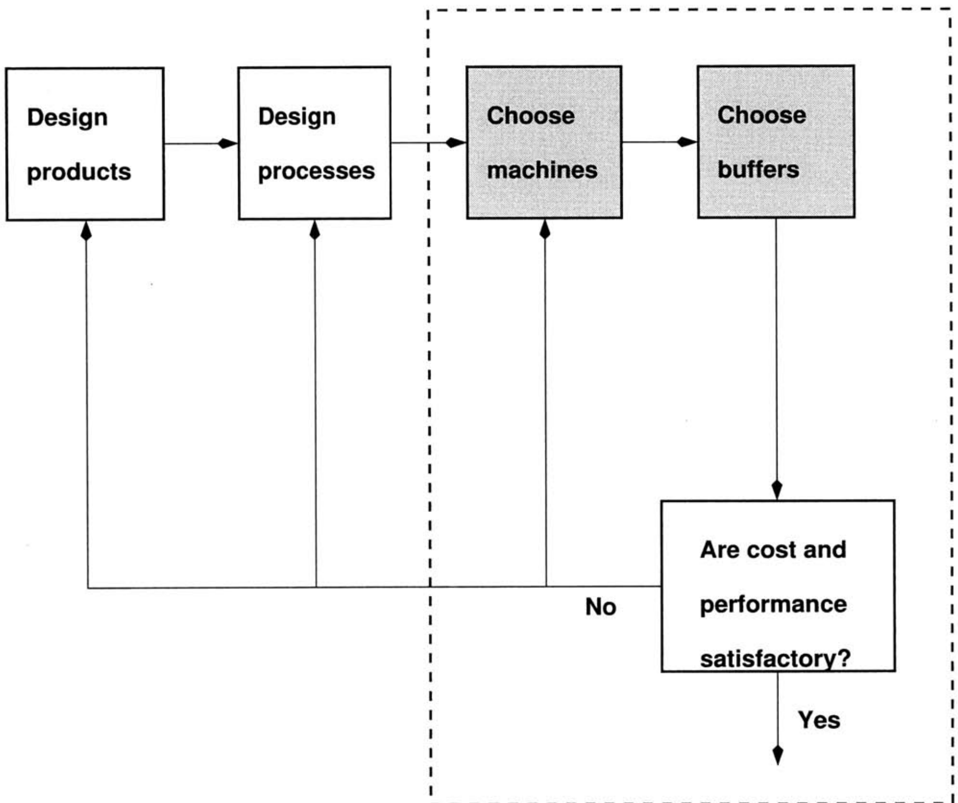


Figure 7-1: Product/process/system design

We consider the problem of machine selection in the design of a flow line. The goal is to choose machines and buffer sizes to maximize the profit generated by the line. We assume that the process has already been designed. That is, the sequence of operations has been chosen, and we must now select the machines to perform the operation at each stage. For some or all of the stages, there are more than one machines available that can perform the operations. These machines may be produced by different vendors, or they may include multiple models produced by the same vendor. The competing machines at each stage have different performance characteristics and different capital costs. We provide a heuristic algorithm to solve this problem. Experimental evidence indicates that it works well.

We use the Buzacott model [Buzacott, 1967a,b, Gershwin, 1994] of a line, in which each machine is characterized by its failure probability (p), its repair probability (r), and, in this problem, its capital cost rate (η). The cost rate η captures the total cost associated with operating the equipment, such as depreciation, electricity, parts, interest and labor. Other characteristics (such as operation time) can be included in extensions of the work reported here. The problem is formulated as an extension of the buffer optimization problems of Gershwin and Schor [2000] and Shi and Gershwin [2009] in which the machines are already specified and the profit rate is a function of production rate, buffer sizes, and average work-in-process inventory. The profit rate function in the new problem is the same as that of the earlier plus a new term for the capital expense rate.

The selection of machines adds an important combinatorial feature to a nonlinear optimization problem. Each combination of machines has a different capital cost and, because of the different reliability characteristics, each combination of machines requires its own optimal set of buffer sizes.

7.1.2 Solution Approach

In previous work on buffer optimization [Gershwin and Schor, 2000, Shi and Gershwin, 2009], buffer sizes were discrete, but the problems were solved efficiently by treating buffer sizes as though they were continuous variables. (When the continuous-variable optimum buffer size vector was found, it was rounded to a nearby vector of integers.) This was possible because all the functions of buffer size that entered into the objectives and the constraints were meaningful when extended to continuous variables. In effect, Gershwin and Schor [2000] embedded the discrete set of possible buffer size vectors in a continuous space, solved the optimization problem, and rounded the solution.

The present problem is also solved by embedding the set of possible machine combinations in a continuous space. For a k-machine line, each combination of machines can be represented by a discrete point in a 3k dimensional space. This is the space of all possible $(r_i, p_i, \eta_i), i = 1, ..., k$. A hyperplane cutting heuristic is developed to solve the discrete combinatorial problem of simultaneously selecting machines and buffers for a flow line to maximizes the profit rate. Through numerical experimentation, the heuristic has been found to be accurate and fast.

7.1.3 Literature Review

This research draws on prior work in unreliable production rate assessments, buffer allocation and machine selection methods for profit maximization.

There have been many techniques developed to evaluate the production rate of unreliable production lines, including Koenigsberg [1959], Buxey et al. [1973], Buzacott and Hanifin [1978], Davis and Stubitz [1987], Hillier and So [1991], Dallery and Gershwin [1992], Buzacott and Shanthikumar [1993], Papadopoulos et al. [1993], Gershwin [1994], Papadopoulos and Heavey [1996], Altiok [1997] Tempelmeier and Burger [2001], Tempelmeier [2003], Li and Meerkov [2009]. We develop the current work based on the deterministic processing time, discrete material model with a single failure mode [Schick and Gershwin, 1978, Buzacott and Shanthikumar, 1993, Gershwin, 1994].

Many authors have written about methods to optimize buffers between unreliable machines or qualitative properties of optimal buffer space allocation including Hillier and Boling [1979], Soyster et al. [1979], Hillier et al. [1993], Jacobs and Meerkov [1995], Han and Park [2002], Sadr and Malhame [2004]. Schor [1995] and Gershwin and Schor [2000] develop four efficient algorithms for optimal buffer allocation. Of particular interest is the maximization of the profit of a production system with no upper limit on buffer capacity. The profit includes costs for buffer capacity and work in progress. This paper extends this work to include machine selection. For machine selection subject to a production rate constraint, we build upon the method developed in Shi and Gershwin [2009].

There is also considerable research literature in machine selection for production system design using various methods. It includes Myint and Tabucanon [1994] (multiple-criteria optimization), Lin and Yang [1996] (the analytical hierarchy process), Beaulieu et al. [1997] (a two-phase heuristic based on grouping machines into cells), Chtourou et al. [2005] (expert systems). More recently, Nahas et al. [2014] developed a Genetic Algorithm for selecting machines and buffers in Assembly/Disassembly networks subject to a cost constraint. Our technique focuses on transfer lines with production rate constraints and differs in its approach and computational efficiency.

In this paper, we develop a hyperplane cutting heuristic method for the machine selection problem. Hyperplane cutting techniques for mixed integer linear programs (MILPs) were first developed in theory by Gomory [1960]. In Gomory's hyperplane cutting method, a cut creates a linear constraint that does not exclude any feasible integer solutions of the problem under consideration. His cuts create additional linear constraints which allow the simplex method to be applied more efficiently. The simplex method is applied to find the optimal solution, and a new cutting plane is applied and the method repeated. Many subsequent methods for linear programming problems are based on this approach. Westerlund and Pettersson [1995] describe an extended cutting plane algorithm to solve convex mixed integer nonlinear problems (MINLPs) with "a moderate degree of nonlinearity." The objective function of this paper is nonlinear, and the method we develop also uses hyperplanes to separate the feasible solutions but approach the problem by reducing the size of the solution space.

7.1.4 Outline

The problem is described precisely in Section 7.2. An algorithm, for a version of the problem in which there is no production rate requirement, is developed intuitively and then stated precisely in Section 7.3. Section 7.4 provides numerical evidence of the algorithm's accuracy and of its computer time requirements for this problem. Then we extend this algorithm to the production-rate-constrained problem in Section 7.5 and describe numerical experience in Section 7.7. Section 7.8 concludes and suggests further related research.

7.2 Problem Description

7.2.1 Technical Overview

We seek to maximize the profit of a production line by choosing the best set of available machines and buffer sizes. The production process has been selected; now we must choose a machine to perform the specified operation at each stage of the line and an appropriate set of in-process inventory buffers. A possible set of machines is a set of machines in which there is one machine to perform the stage i operation for each stage i. We select the optimal set of machines, M^*, from the set of all possible sets of machines $\mathcal{M}$. Simultaneously, we select the set of optimal buffer capacities between machines, N^*. Machine choices are illustrated in Figure 7-2.

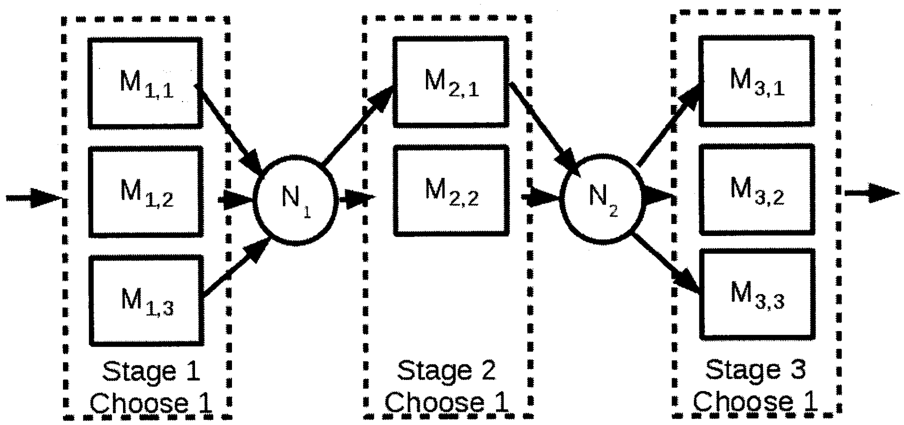


Figure 7-2: Choices for machines at each stage of a three-stage line

The production line has k stages and there are s_i choices for a machine at stage i. There are therefore $S = \prod_i^k s_i$ possible ways to select a machine set. We denote choice j at stage i by M_{ij}. A selection of a set of machines is denoted $M = (M_{1j_1}, M_{2j_2}, ...M_{kj_k})$. We assume a Buzacott model [Buzacott, 1967a,b, Gershwin, 1994] in which all operation times are equal to one time unit, and machines are unreliable. The times to fail and to repair are geometrically distributed. The reliability parameters of machine M_{ij} are p_{ij}, the probability of a failure during a time unit while the machine is operating, and r_{ij}, the probability of a repair during a time unit while the machine is down. In addition, machines have a parameter η_{ij}, the fully burdened cost per time unit of owning and operating the machine. Transportation time is negligible compared to the operation time. Figure 7-3 shows the parameters.

Since each M_{ij} has the parameters p_{ij}, r_{ij}, and η_{ij}, every M has a corresponding parameter set x^M with $3k$ elements which is organized according to $x^M_{3i-2,j} = p_{ij}$, $x^M_{3i-1,j} = r_{ij}$, $x^M_{3i,j} = \eta_{ij}$ for $j = 1, ..., s_i$ and $i = 1, ..., k$, where j refers to the jth choice of machine at stage i.

Consider again the three stage production line in figure 7-2. In this system there are three stages,

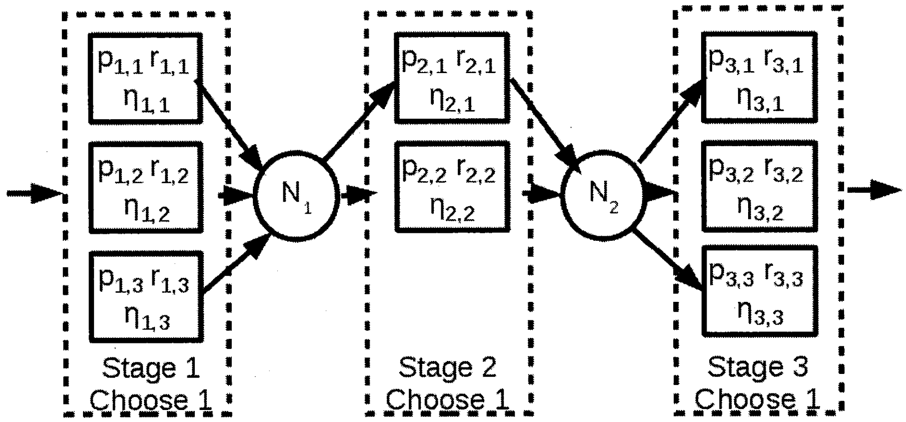


Figure 7-3: Parameters p_{ij}, r_{ij}, and η_{ij} for each machine M_{ij}

so $k = 3$. The first stage has three choices so $s_1 = 3$, and likewise $s_2 = 2$. If we were interested in a particular machine set that included the second machine in the first stage, the second machine in the second stage and the first machine in the third stage we would denote it $M = (M_{1,2}, M_{2,2}, M_{3,1})$ and its parameter vector, x_1^M would be $(p_{1,2}, r_{1,2}, \eta_{1,2}, p_{2,2}, r_{2,2}, \eta_{2,2}, p_{3,1}, r_{3,1}, \eta_{3,1})$.

The problem and solution described here are built on the buffer optimization methods of Gershwin and Schor [2000] and Shi and Gershwin [2009]. We make all the assumptions and approximations of those models, and use similar notation. We outline the key features of the model below.

In selecting unreliable machines, one must also consider the effect of storing work in progress in buffers. The capacity of buffer B_i is N_i, and the steady state average number of items in B_i is $\bar{n}_i$. c_i is the cost per time unit per unit of buffer capacity, and b_i is the cost per time unit per item for storing work in progress in B_i. The production rate P and the average inventory $\bar{n}$ of a particular line M are nonlinear functions of x and N. For a given M and N we calculate P and $\bar{n}$ according to the decomposition in Gershwin [1994].

The decision variables in this model are M and N. Given M, the set of buffers N that maximizes profit is determined according to Shi and Gershwin [2009]. They determine the optimal allocation of N (denoted N^*) which optimizes the profit (Π_1), which is given by

$$\Pi_1(N) = AP - \sum_{i=1}^{k-1} b_i \bar{n}_i - \sum_{i=1}^{k-1} c_i N_i$$

This profit rate function is divided into three parts. The first term, AP, is the revenue rate term. It captures the amount of money per unit time that the production system generates as a result of producing the final product. The second term captures the cost of holding work in progress. Shrinkage, damage, and exposure to obsolescence are some factors that a system designer might consider when determining the cost parameter b to apply to the inventory term $\bar{n}$. The final term is the cost of factory floor space, which is required for holding buffer parts.

This thesis and the work of Shi and Gershwin [2009] also includes a provision for a production rate constraint, $P \geq P^*$. Many factory systems have a minimum production target that is set exogenously by customer demands, managerial goals, or sales expectations. The P^* constraint captures this common business practice.

We extend this profit function with the inclusion of η as shown below in (7.1).

7.2.2 Technical Problem Statement

Production-rate-constrained problem

In the form of the problem considered here, we maximize the profit rate (7.1) subject to constraints on the minimum sizes of buffers and the minimum acceptable production rate. We include a new term, η, which represents to total operating cost per unit time of a machine. Machines vary in operating and capital costs, and this term captures these additional expenses.

Problem: Choose $M = (M_{1j_1}, M_{2j_2}, \ldots, M_{kj_k}) \in \mathcal{M}$ and $N = (N_1, N_2, ..., N_{k-1})$ to maximize

$$\Pi_2(M,N) = AP - \sum_{i=1}^{k-1} b_i\bar{n}_i - \sum_{i=1}^{k-1} c_i N_i - \sum_{i=1}^{k} \eta_i \tag{7.1}$$

subject to

$$P > P^* \tag{7.2}$$

$$N_i \geq N_{\text{min}}$$

This is a combinatorial optimization because $\mathcal{M}$ is a set of combinations of discrete choices.

Production-rate-unconstrained problem

This version of the problem is the same as the previous except that we drop (7.2).

Problem: Choose $M = (M_{1j_1}, M_{2j_2}, \ldots, M_{kj_k}) \in \mathcal{M}$ and $N = (N_1, N_2, ..., N_{k-1})$ to

maximize

$$\Pi_2(M,N) = AP - \sum_{i=1}^{k-1} b_i\bar{n}_i - \sum_{i=1}^{k-1} c_i N_i - \sum_{i=1}^{k} \eta_i \tag{7.3}$$

subject to

$$N_i \geq N_{\text{min}}$$

We describe an algorithm to solve the production-rate-unconstrained problem in Section 7.3.

Numerical experiments and examples are presented in Section 7.4. Then we extend this algorithm to the production-rate-constrained problem in Section 7.5 and describe numerical experience in Section 7.7.

7.3 Algorithm for the Unconstrained Problem

Here, we develop a cutting plane heuristic to find the optimal, or near optimal M^* which maximizes the profit function Π_2.

7.3.1 Notation and Assumption

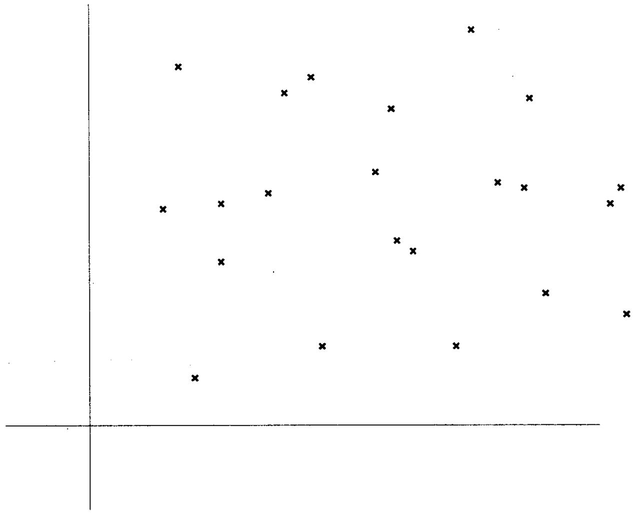
It is convenient to define $\hat{\Pi}_2(x, N)$, $N^*(x)$, and $\Pi_2^*(x)$. If we define $p_{ij} = x_{3i-2,j}, r_{ij} = x_{3i-1,j}$, and $\eta_{ij} = x_{3i,j}$, then $\hat{\Pi}_2(x, N)$ is the profit function that can be evaluated at all values of x that satisfy $0 \leq x_{3i-2,j} < 1$, $0 \leq x_{3i-1,j} < 1$, $x_{3i,j} \geq 0$ for $i = 1, ..., k$. By doing this, we are embedding the large discrete finite set of possible choices into a much larger continuous space. $N^*(x)$ is the vector of buffer sizes that maximizes $\hat{\Pi}_2(x, N)$ for a given x, and $\Pi_2^*(x) = \hat{\Pi}_2(x, N^*(x))$, the value of $\hat{\Pi}_2(x, N)$ when $N = N^*(x)$. In the following, we make the important assumption that $\Pi_2^*(x)$ is a smooth, convex function of x.

It is also convenient to define x^M as the value of x that corresponds to machine set M.

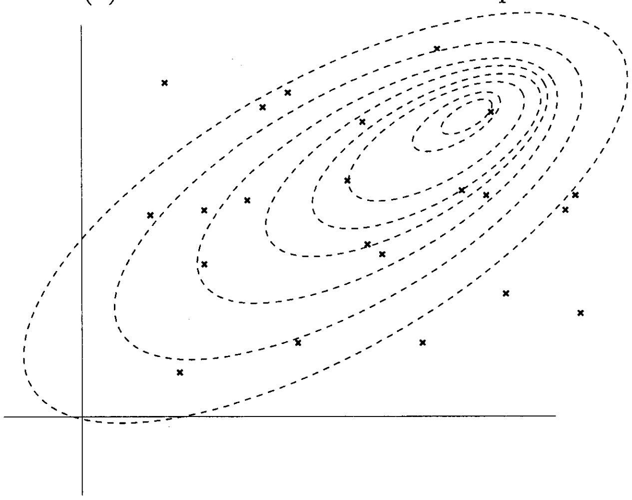
7.3.2 Overview of Algorithm

We represent the $3k$-dimensional space of all x in Figure 7-4a. Each point indicated by **X** represents an x^M, the vector of parameters of machine set M. The strategy of this heuristic iterative algorithm is to eliminate close to half the points at each iteration. We construct a plane that separates the least likely candidates for optimality from the most likely. This is repeated until only one candidate remains.

Figure 7-4b shows the contours of constant $\Pi_2^*(x)$. The smaller the contour, the larger $\Pi_2^*(x)$. The goal is to find the point x^M that lies on the smallest contour. Because there are many points, we must do calculations that involve as few points as possible, and any calculation that does involve



(a) Parameter vectors x^M at Step 0.



(b) Iso-profit rate contours of Π_2.

Figure 7-4: Orientation to Parameter Space

a large number of points must be very simple. (Note that we do not need to calculate the contours. They are shown for intuitive visualization only.)

It is reasonable to assume that a good cutting plane will pass close to the average x^M. Therefore, we calculate the average:

$$\bar{x} = \frac{1}{S} \sum_M x^M$$

The sum is over all $M \in \mathcal{M}$. There are S terms in the sum. This can also be written, for stage i,

$$\bar{r}_i = \frac{1}{s_i}\sum_{j=1}^{s_i} r_{ij}; \qquad \bar{p}_i = \frac{1}{s_i}\sum_{j=1}^{s_i} p_{ij}; \qquad \bar{\eta}_i = \frac{1}{s_i}\sum_{j=1}^{s_i} \eta_{ij} \tag{7.4}$$

where $p_{ij} = x_{3i-2,j}, r_{ij} = x_{3i-1,j}$, and $\eta_{ij} = x_{3i,j}$ for all $j = 1, ..., s_i; i = 1, ..., k$.

The average is the new point indicated in Figure 7-5a. To construct a good plane, we calculate the gradient of $\Pi_2^*(x)$ at $\bar{x}$, which is shown in Figure 7-5b. We draw the plane orthogonal to the gradient, and we mark for elimination all the points on the side of the plane opposite the gradient direction.

The surviving points are shown in Figure 7-6. We now repeat the process until all points but one are eliminated. The remaining point is the estimate of the optimum that the algorithm provides. The corresponding M is the set of machines and the last N is the corresponding vector of buffer sizes. A formal statement of the algorithm follows.

7.3.3 Algorithm Statement

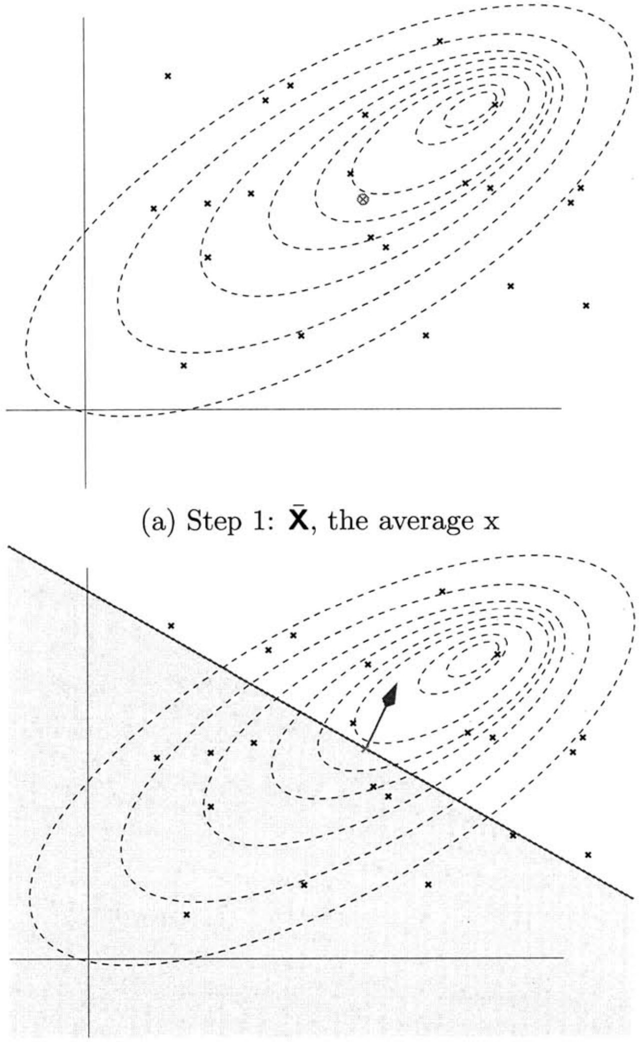
- **Step 0:** Define $\mathcal{M}^0 = \mathcal{M}$. Set $I = 0$.

- **Step 1:** Calculate $\bar{x}^I$ according to (7.4) where the sum is over all x corresponding to M, $M \in \mathcal{M}^I$.

- **Step 2:** Calculate the approximate gradient G_i^I at $x = \bar{x}^I$ according to

$$G_i^I = \frac{\Pi_2^*(x_1, ..., x_i + \epsilon, ..., x_{3k}) - \Pi_2^*(x_1, ..., x_i, ..., x_{3k})}{\epsilon}, i = 1, ..., 3k \tag{7.5}$$

- **Step 3:** Construct $\mathcal{M}^{I+1}$ as follows:



(a) Step 1: $\bar{\mathbf{X}}$, the average x

(b) Step 2: Construction of the cutting hyperplane

Figure 7-5: Location and Construction of the Cutting Surface

$$\mathcal{M}^{I+1} = \{M \in \mathcal{M}^I | G^I(x - \bar{x}^I) > 0 \text{ for } x \text{ corresponding to } M\} \tag{7.6}$$

- **Step 4:** If $\mathcal{M}^{I+1}$ consists of a single machine set M, stop. M is the result. Otherwise, increment I and go to Step 1.

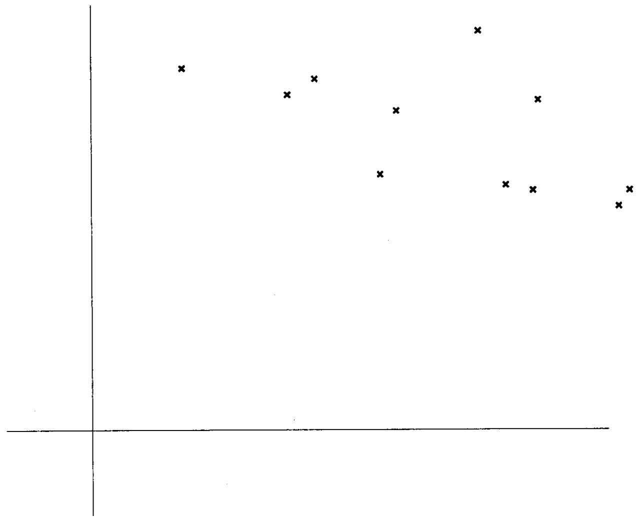


Figure 7-6: Step 3: Elimination of points below the cutting hyperplane

7.4 Numerical Experiments

In this section, numerical experiments are performed to determine the accuracy and performance characteristics of the algorithm. In many of the experiments that follow, machine parameters were generated with pseudo-random numbers. Parameters r_i, p_i, and η_i were uniformly distributed. The revenue coefficient A and the cost coefficients b_i and c_i were fixed. Table 7.1 provides a summary of the parameters:

Parameter	Value or Range
p_i	[.0085,.0185]
r_i	[.095,.145]
η_i	[1.25,10.25]
A	2000
b	(1,1,1)
c	(1,1,1)

Table 7.1: Data for the random machine cases

7.4.1 Short Lines

To test the heuristic, we ran it on 10,000 four-stage lines ($k = 4$) in which there were four choices for each stage ($s_i = 4, i = 1, 2, 3, 4$).

Line Length (k)	Accuracy, %	Precision, %
7	95	.2
8	95	.4
9	100	100

Table 7.2: Results of the long line cases

We compared the optimum found by the algorithm with that found by enumeration. We found that the results of the heuristic agreed with the enumeration results in 98.2% of the cases. We also found that for the cases in which the heuristic did not get the correct solution, the average error in the profit rate was 1.2%. (The average over all cases was much smaller.) The average computation time for the heuristic was 35.2 seconds; the average time for enumeration was 3.54 minutes.

Special case Consider a line such that for some stage i, there is a machine which is better in all respects than all the other machines available for that stage. That is, there exists some $j' \in \{1, ..., s_i\}$ such that $p_{j'} < p_j$; $r_{j'} > r_j$; $\eta_{j'} < \eta_j$ for all $j = 1, ..., j' - 1, j' + 1, ..., s_i$. $M_{ij'}$ fails less often, takes less time to repair, *and* costs less than any other machine that is available for stage i. We would expect that such a machine would always be included in the optimal design.

We experimented with four-stage systems. The solution always agreed with our expectation that if there is a machine available that out-performs all the other choices in the same stage, it will always be chosen. In addition, the solution found by the plane-cutting heuristic always agreed with the solution obtained by enumeration.

7.4.2 Long Lines

Initial trials on longer lines were conducted on systems with up to ten stages with three choices at each stage. Verification by complete enumeration would take a prohibitive amount of time, so accuracy was tested by comparing the profit rate of the heuristic solution M^* with the profit rates of a small set of neighboring M. Parameters were chosen using a pseudo-random number generator, and results are summarized in Table 7.2. Accuracy means the percentage of cases that got the exact result (when compared with neighbors), and precision is the average percentage error in profit rate, considering only the cases in which the heuristic did not get the exact results.

The advantage of the plane cutting technique becomes apparent in its computational time. For cases in which the line has k stages and there are s machine choices for each stage, calculating the profit rate of every possible machine set requires s^k evaluations of the profit rate function, and these evaluations include buffer optimization. The time required for this operation grows exponentially with the number of stages [Shi and Gershwin, 2009], a method which quickly becomes impractical. Hyperplane cutting reduces the number of these evaluations to approximately $(3k+1)\log_2(s)$. The $\log_2 s$ term captures the number of iterations required if each iteration eliminates approximately half of the decision space. The $3k+1$ term describes the number of times the gradient is calculated for each iteration. Each time the gradient is estimated, the profit rate is calculated so the running time must scale directly with this quantity. Figure 7-7 shows estimates of the computer time required for both methods.

Line Length (k)	Enumeration	Heuristic
4	81	24
5	243	27.86
6	729	31.02
7	2187	33.69
8	6561	38.04
9	19683	39.86
10	59049	41.51

Table 7.3: Estimates of algorithm loops

7.4.3 Example Problem

This section presents an example solution to the design of a particular manufacturing system. In the following example problem, there are 4 stages, with 4 choices per stage. All parameters for each machine were generated randomly, yielding a set of parameters presented in Table 7.4. The

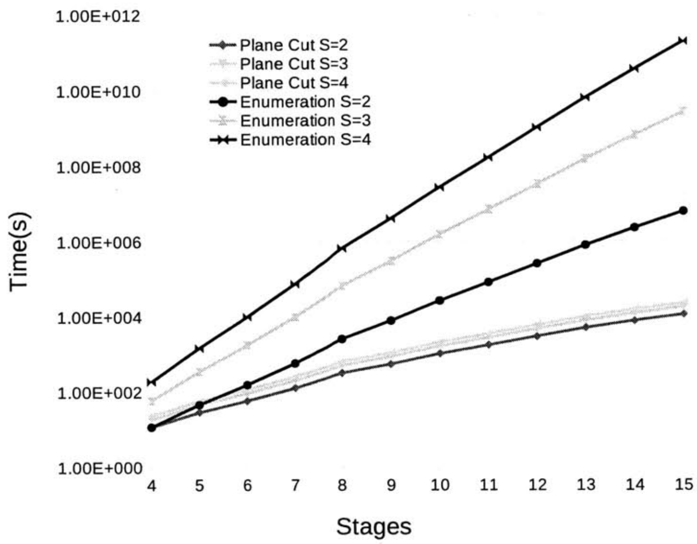


Figure 7-7: Estimates of Computer Time: Enumeration Vs Hyperplane Cutting

production rate is unconstrained. There are $4^4 = 256$ possible machine sets and the dimension of the parameter vector space is $4 \times 3 = 12$.

For this small problem, we enumerate all 256 possible machine sets and solve the buffer design problem for each. Solving the buffer design problem is the most time-consuming step and complete enumeration is impractical for large systems, but it is useful to visualize the progress and accuracy of the algorithm when direct comparison by enumeration is available. First, the profit for each machine set is solved by enumeration. Next, the profits are order from most profitable to least, and plotted in Figure 7-8a. Figure 7-8 and Figure 7-9 show how the enumerated outcomes are eliminated at each iteration of the algorithm. Nearly half of the possibilities are eliminated at each step, and this case converges to the optimal solution.

Stage 1

p	r	η	e
.013795	.10946	1.7636	.8881
.01128	.099593	2.2373	.8983
.011056	.10562	1.5822	.9052
.0095437	.10208	1.5366	.9145

Stage 2

p	r	η	e
.012036	.11412	1.4343	.9046
.016487	.11228	2.0593	.8720
.010163	.11265	2.4684	.9172
.013803	.11046	2.006	.8889

Stage 3

p	r	η	e
.017135	.10987	2.5511	.8651
.017524	.096579	1.5677	.8464
.015596	.097038	1.7309	.8615
.014682	.11093	1.7075	.8831

Stage 4

p	r	η	e
.0085538	.10903	1.6119	.9273
.016997	.10687	1.8329	.8628
.0094287	.11401	2.0331	.9236
.014126	.099772	2.3997	.8760

Table 7.4: Parameters for the unconstrained example problem

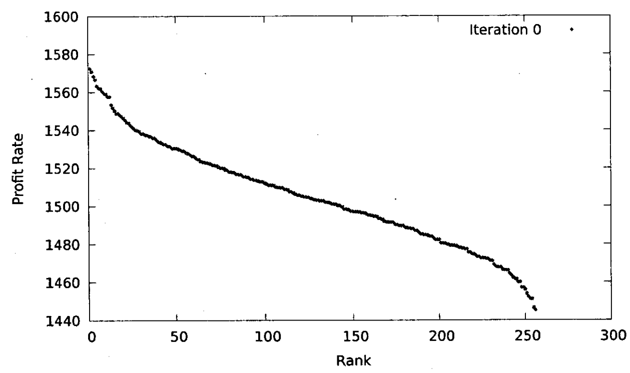
7.5 Problems with the Production Rate Constraint

We next consider the original problem statement in section 7.2.2, where the system must be designed to achieve a nonzero minimum production rate, $P \geq P^* > 0$.

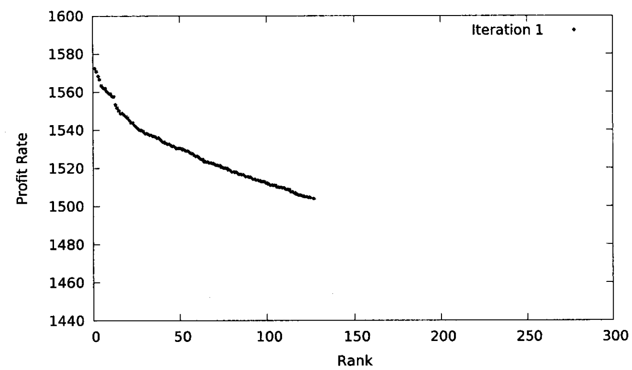
We proceed by modifying the algorithm in 7.3.3 with the inclusion of filtering step after the feasible region is defined.

- **Step 0.5:** filter out all machine sets M such that if any machine $M_{ij} \subset M$ has $e_{ij} \leq P^* + \epsilon$.

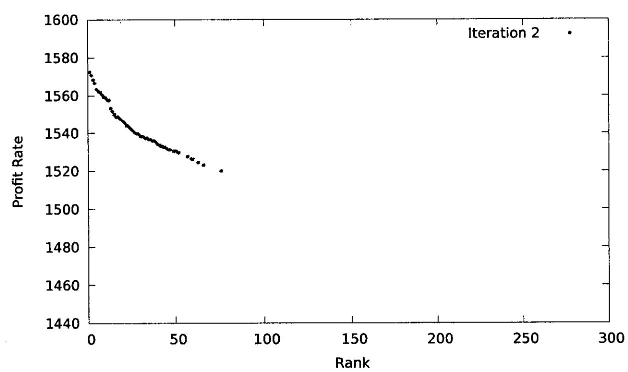
This filter ensures that the remaining M are feasible solutions, as a machine with $e < P*$ would never be able to produce quickly enough to meet the requirements. The inclusion of the small offset ϵ is necessary to ensure convergence, as production systems with $e_{ij} = P^*$ would require infinite buffers. However, this is not an important restriction as no line designer for a stochastic environment with a target production rate of P^* would consider a machine with $e = P^*$.



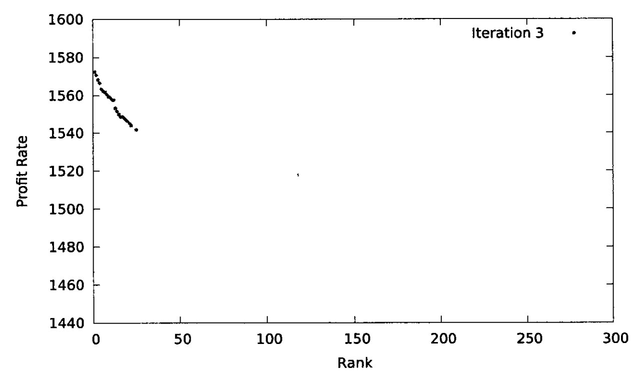
(a) Initial Set



(b) Iteration 1



(c) Iteration 2



(d) Iteration 3

Figure 7-8: Illustration of reduction of feasible solutions for solution to 7.4, Part 1

7.6 Numerical Experiments with the Production Rate Constraint

We repeat the random machine experiments from Section 7.4 of with the same parameters as before, as in Table 7.1 but now we add constraint 7.2. So that the problem is meaningful, we choose P^* to satisfy:

$$P^* = \min_j \frac{1}{N_j} \sum_i e_{i,j} \tag{7.7}$$

This specification of P^* guarantees that the minimum production rate is in the middle of the isolated efficiencies for the available machines. This is important because a P^* which is too small would be the same as the cases in Section 7.4, and a P^* which is too large would eliminate all of the available machines and result in zero possible solutions.

10,000 trials were run for $k = 4$ stage lines and 1000 trials were run for $k = 10$ stage lines. The results of these experiments are shown in Table 7.5. They show remarkable accuracy and precision.

Line Length (k)	Accuracy, %	Precision, %
4	94	.25
10	94	.5

Table 7.5: Results of experiments with production rate constrains

7.7 Numerical Experience

In this section we explore two case studies which further develop the important insights and limitations that can be garnered from this method of investigation. First, we explicitly explore the trade-off between reliability and cost. We conclude this section is an investigation of situation which will cause the heuristic to become inaccurate.

7.7.1 The Balance Between Cost and Reliability

In this section we conduct an experiment that investigates when it is worthwhile to invest in machines that are more expensive, and more reliable, and when it is better to buy inexpensive machine and compensate for the inefficiency with extra buffer space.

Table 7.6 shows identical machines except for a scaling factor applied to the cost rate η in all choices of machines in the second stage. γ ranges from 0 to 50. For small γ, all machines are relatively cheap. For large γ, reliable machines become much more expensive than less reliable machines.

The machines that were selected in stage 2 are shown in Table 7.7 for each value of γ. The figure shows that for $\gamma < 10$, the most expensive, most reliable machine in stage 2 is chosen. This reflects a classical lean manufacturing environment where buffers are small and machines are reliable. At the other extreme, for $\gamma > 35$, reliable processes are very expensive. The profit of the manufacturing system is optimized by selecting a low cost, low efficiency machine and compensating for machine breakdowns with buffer capacity.

Stage 1

p	r	η	e
.013	.095	100	.87
.011	.099	150	.90
.0095	.105	200	.917
.009	.11	250	.92

Stage 2

p	r	η	e
.013	.095	100γ	.87
.011	.099	150γ	.90
.0095	.105	200γ	.082
.009	.11	250γ	.076

Stage 3

p	r	η	e
.013	.095	100	.87
.011	.099	150	.90
.0095	.105	200	.082
.009	.11	250	.076

Stage 4

p	r	η	e
.013	.095	100	.87
.011	.099	150	.90
.0095	.105	200	.082
.009	.11	250	.076

Table 7.6: Parameters for the example

γ Range	Stage 2 Machine %
1-10	$M_{2,4}$
11-25	$M_{2,3}$
26-35	$M_{2,2}$
36-50	$M_{2,1}$

Table 7.7: Effect of varying γ

Figure 7-10 shows the piecewise smooth transition between stage 2 machines as the scaling constant is increased. While the profit function is continuous, the total buffer size in Figure 7-10 is discontinuous. At each discontinuity, a different machine in stage 2 becomes optimal, and the buffers are adjusted to maximize the profit of the new system. The figure tells a story of very different factory operation strategies. Factories with unreliable machines pay to reach their production goals with inventory and floorspace implying very different management and workplace environments and exposure to risks associated with high levels of work in progress.

7.7.2 Limitations

This thesis presents a heuristic for nonlinear mixed integer program, and all heuristics (as well as all other known solutions for similar problems) have limits. The assumptions that the profit function is smooth and convex is imposed and the computation of the profit gradient may be inaccurate under special conditions.

The underlying productivity model is based on the work of [Gershwin, 1994]. As P^* approaches e this model becomes inaccurate due to the very large size of buffers.

To illustrate this effect of this breakdown we create a set of machines with properties in table 7.8 and hold them as the only available choice. Next, we vary P^* between .79 and .86 with a .002 step size and plot the buffer sizes that result in Figure 7-11.

Stage 1

p	r	η	e
0.014515	0.11128	1.5697	0.8846

Stage 2

p	r	η	e
0.017153	0.1047	2.855	0.85923

Stage 3

p	r	η	e
0.01373	0.102	2.3336	0.8814

Stage 4

p	r	η	e
0.015696	0.09773	2.807	0.8616

Table 7.8: Parameters for the example

Machine designers must be careful to avoid using this algorithm for systems which are are barely efficient enough to meet their production rate constraints.

This challenge to algorithm accuracy is easily mitigated with two approaches;

1. Filter out machines for which e is only marginally above P^*. This approach is outlined in Section 6.5.

2. As the algorithm approaches the final iterations, switch to a enumerating all remaining possible solutions.

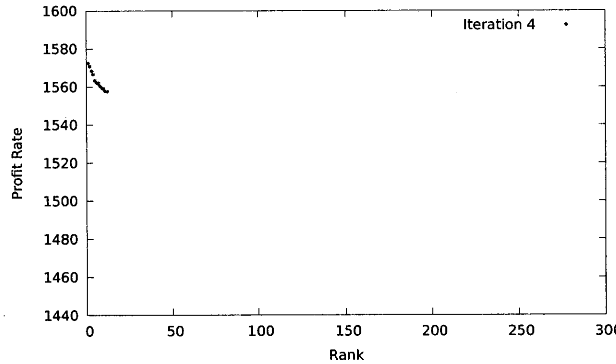
7.8 Conclusion

7.8.1 Summary

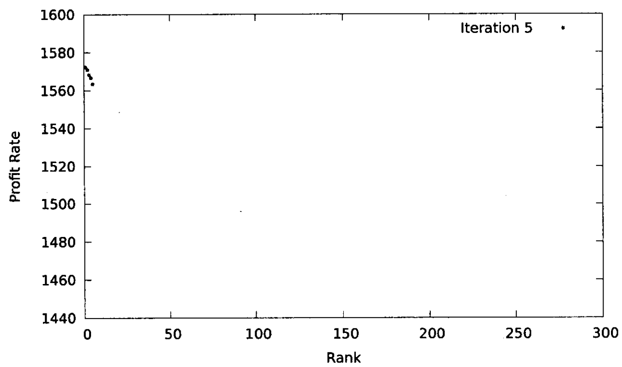
A fast, accurate, flexible heuristic has been developed to make optimal, or near-optimal decisions about selecting unreliable machines in a factory setting. We have reduced a problem which would take years to solve by enumeration to one which converges in minutes with a high degree of accuracy.

7.8.2 Future Work

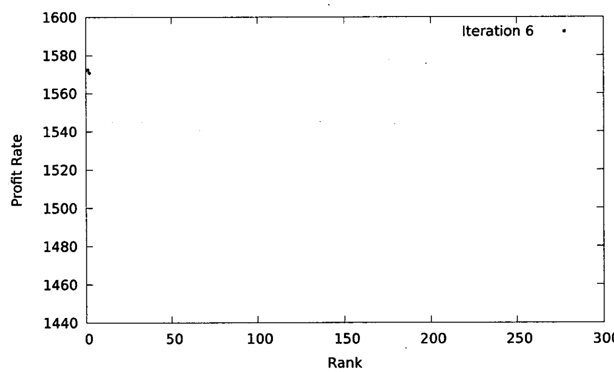
Gains in accuracy and time may be attainable by switching to enumeration when there are few choices left in $\mathcal{M}^{\mathcal{I}}$. Useful extensions include allowing more general configurations of machines such as machines in parallel at a stage. Other models of machines may provide additional design options. For example, machines may be available that operate at different speeds. This would require a fourth parameter in addition to those described here and a different model for productivity. More general configurations and models, such as multiple failure modes, assembly/disassembly systems, and systems with loops may also be treatable by this solution technique. The cost of buffers can include a cost of having a buffer, independent of its size or amount of inventory. This could represent the cost of the material handling mechanism for buffers. In that case, the optimal solution may have many fewer buffers than machines. Other extensions may include production rate, a total inventory space constraints or a capital expense constraint.



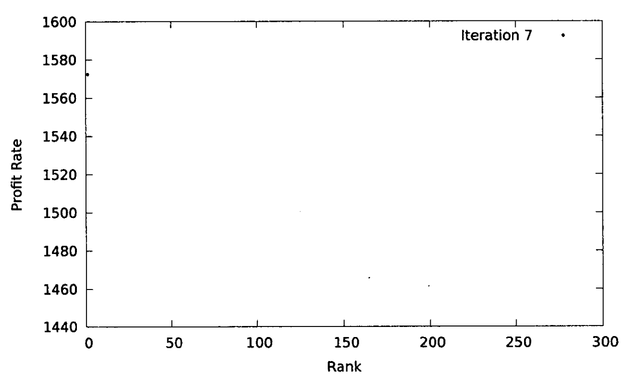
(a) Iteration 4



(b) Iteration 5



(c) Iteration 6



(d) Iteration 7

Figure 7-9: Illustration of reduction of feasible solutions for solution to 7.4, Part 2

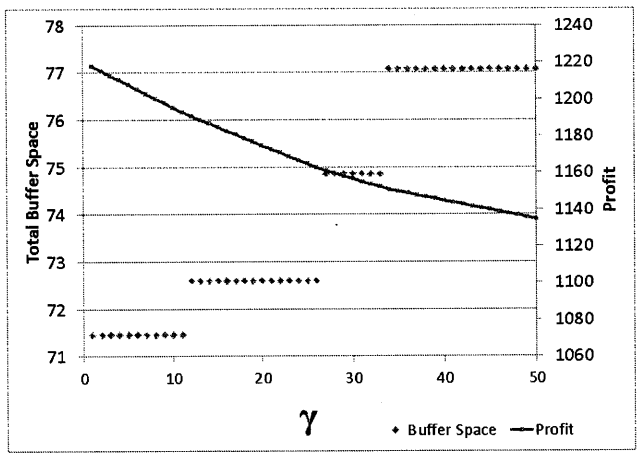


Figure 7-10: Buffers can smooth profitability

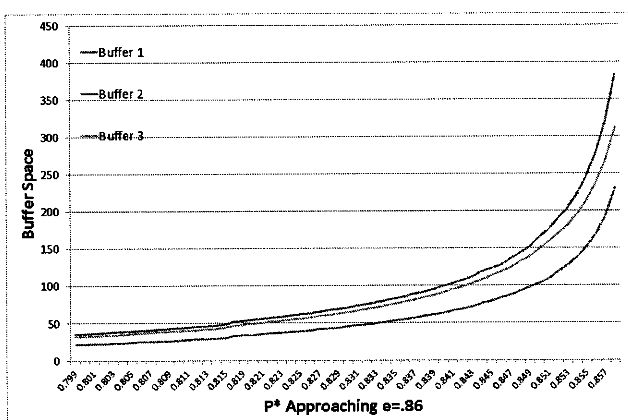


Figure 7-11: Buffers can grow rapidly when P^* is close to e

Appendix A

	Normal Staffing All	Peak TISS	SOFA	TISS	Upper Vents	First 24H
count	8724.000000	8724.000000	8724.000000	8724.00000	8724	8724.000000
mean	0.848630	0.231712	11.659270	21.71904	0.2077029	0.445104
std	0.292579	0.200291	3.216726	5.47100	0.4056862	0.255453
min	0.020833	0.000000	0.000000	0.00000	0	0.000000
25%	0.640625	0.000000	10.111111	18.00000	0	0.285714
50%	0.833333	0.200000	11.888889	21.00000	0	0.428571
75%	1.031702	0.333333	13.700000	24.50000	0	0.600000
max	2.000000	1.500000	23.000000	81.00000	1	3.000000

	Admits	New Nurse All	EU Critical All	Float Nurse All	Discharges	Boarders
count	8724.000000	8724.000000	8724.000000	8724.000000	8724.000000	8724.000000
mean	0.154171	0.165912	0.105634	0.227195	0.151572	0.864363
std	0.165534	0.234221	0.172050	0.296836	0.173496	0.159874
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.000000	0.000000	0.000000	0.750000
50%	0.142857	0.000000	0.000000	0.000000	0.125000	0.888889
75%	0.250000	0.333333	0.166667	0.400000	0.250000	1.000000
max	1.000000	3.000000	1.000000	2.666667	1.000000	1.166667

Bibliography

Tayfur Altiok. *Performance Analysis of Manufacturing Systems.* Springer Verlag, New York, 1997.

Derek C Angus. Grappling with intensive care unit quality-does the readmission rate tell us anything? *Critical care medicine*, 26(11):1779–1780, 1998.

A Beaulieu, A Gharbi, and Ait-Kadi. An algorithm for the cell formation and the machine selection problems in the design of a cellular manufacturing system. *International journal of production research*, 35(7):1857–1874, 1997.

Richard Bellman, Richard Ernest Bellman, Richard Ernest Bellman, and Richard Ernest Bellman. *Adaptive control processes: a guided tour*, volume 4. Princeton university press Princeton, 1961.

Sean M Berenholtz, Todd Dorman, Koni Ngo, and Peter J Pronovost. Qualitative review of intensive care unit quality indicators. *Journal of critical care*, 17(1):1–12, 2002.

Leo Breiman, Jerome Friedman, Charles J Stone, and Richard A Olshen. *Classification and regression trees.* CRC press, 1984.

G. M. Buxey, N. D. Slack, and Ray Wild. Production flow line system design — a review. *AIIE Transactions*, 5(1):37–48, 1973.

J. A. Buzacott and L. E. Hanifin. Models of automatic transfer lines with inventory banks — a review and comparison. *AIIE Transactions*, 10(2):197–207, 1978.

John A. Buzacott. Automatic transfer lines with buffer stocks. *International Journal of Production Research*, 5(3):183–200, 1967a.

John A. Buzacott. *Markov chain analysis of automatic transfer lines with buffer stock.* PhD thesis, University of Birmingham, 1967b.

John A. Buzacott and J. George Shanthikumar. *Stochastic Models of Manufacturing Systems.* Prentice Hall, Englewood Cliffs, NJ, 1993.

Pascale Carayon and Ayşe P. Gürses. A human factors engineering conceptual framework of nursing workload and patient safety in intensive care units. *Intensive and Critical Care Nursing*, 21(5):284–301, 2005. ISSN 0964-3397. doi: 10.1016/j.iccn.2004.12.003. URL `http://www.sciencedirect.com/science/article/pii/S0964339705000182`.

Liddy M Chen, Claudio M Martin, Sean P Keenan, and William J Sibbald. Patients readmitted to the intensive care unit during the same hospitalization: clinical features and outcomes. *Critical care medicine*, 26(11):1834–1841, 1998.

Hédi Chtourou, Wassim Masmoudi, and Aref Maalej. An expert system for manufacturing systems machine selection. *Expert Systems with Applications*, 28(3):461–467, 2005.

David C Classen, Roger Resar, Frances Griffin, Frank Federico, Terri Frankel, Nancy Kimmel, John C Whittington, Allan Frankel, Andrew Seger, and Brent C James. 'global trigger tool'shows that adverse events in hospitals may be ten times greater than previously measured. *Health Affairs*, 30(4):581–589, 2011.

Yves Dallery and Stanley B Gershwin. Manufacturing flow line systems: a review of models and analytical results. *Queueing Systems*, 12(1):3–94, 1992.

W. J. Davis and S. J. Stubitz. Configuring a kanban system using a discrete optimization of multiple stochastic responses. *International Journal of Production Research*, 25(5):721–740, 1987.

Maartje De Vos, Wilco Graafmans, Els Keesman, Gert Westert, and Peter HJ van der Voort. Quality measurement at intensive care units: which indicators should we use? *Journal of critical care*, 22(4):267–274, 2007.

S. B. Gershwin. *Manufacturing Systems Engineering.* Prentice-Hall, 1994. Currently available at `http://home.comcast.net/~hierarchy/MSE/mse.html`. For corrections, see `http://web.mit.edu/manuf-sys/www/gershwin.errata.html`.

Stanley B. Gershwin and James E. Schor. Efficient algorithms for buffer space allocation. *Annals of Operations Research*, 93:117–144, 2000.

Ralph Gomory. An algorithm for the mixed integer problem. Technical report, DTIC Document, 1960.

FA Griffin and RK Resar. Ihi global trigger tool for measuring adverse events. *Institute for Healthcare Improvement Innovation Series White Paper*, 2009.

Man-Soo Han and Dong-Jo Park. Performance analysis and optimization of cyclic production lines. *IIE Transactions*, 34(4):411–422, 2002.

Seetharaman. Hariharan, Deryk. Chen, Lorna. Merritt-Charles, Nahmorah. Bobb, Loren. DeFreitas, Joann Mohamed. Esdelle-Thomas, Delise. Charles, Karen. Colley, and Elise. Renaud. The utilities of the therapeutic intervention scoring system (TISS-28). *Indian Journal of Critical Care Medicine*, 11(2):61–66, 2007. doi: 10.4103/0972-5229.33387.

Alex B Haynes, Thomas G Weiser, William R Berry, Stuart R Lipsitz, Abdel-Hadi S Breizat, E Patchen Dellinger, Teodoro Herbosa, Sudhir Joseph, Pascience L Kibatala, Marie Carmela M Lapitan, et al. A surgical safety checklist to reduce morbidity and mortality in a global population. *New England Journal of Medicine*, 360(5):491–499, 2009.

F. S. Hillier and K. So. The effect of machine breakdowns and interstage storage on the performance of production line systems. *International Journal of Production Research*, 29(10):2043–2055, 1991.

Frederick S. Hillier and Ronald W. Boling. On the optimal allocation of work in symmetrically unbalanced production line systems with variable operation times. *Management Science*, 25(8): 721–728, 1979.

F.S. Hillier, K.C. So, and R.W. Boling. Toward characterizing the optimal allocation of storage space in production line systems with variable processing times. *Management Science*, 39(1): 126–133, 1993.

D. A. Jacobs and S. M. Meerkov. A system-theoretic property of serial production lines: Improvability. *International Journal of Systems Science*, 26:755–785, 1995.

U Janssens, C Graf, J Graf, PW Radke, B Königs, K Ch Koch, W Lepper, J Vom Dahl, and P Hanrath. Evaluation of the sofa score: a single-center experience of a medical intensive care unit in 303 consecutive patients with predominantly cardiovascular disorders. *Intensive care medicine*, 26(8):1037–1045, 2000.

A Roberta Keene and David J Cullen. Therapeutic intervention scoring system: update 1983. *Critical care medicine*, 11(1):1–3, 1983.

Kenji Kira and Larry A Rendell. A practical approach to feature selection. In *Proceedings of the ninth international workshop on Machine learning*, pages 249–256, 1992.

Ernest Koenigsberg. Production lines and internal storage — a review. *Management Science*, 5(4): 410–433, 1959.

Lucian L Leape. Error in medicine. *Jama*, 272(23):1851–1857, 1994.

Nancy Leveson. *Engineering a safer world: Systems thinking applied to safety.* Mit Press, 2011.

Jingshan Li and Semyon M. Meerkov. *Production Systems Engineering.* Springer, New York, NY, USA, 2009.

Zone-Ching Lin and Chu-Been Yang. Evaluation of machine selection by the AHP method. *Journal of Materials Processing Technology*, 57(3):253–258, 1996.

Elizabeth A Martinez, Karen Donelan, Justin P Henneman, Sean M Berenholtz, Paola D Miralles, Allison E Krug, Lisa I Iezzoni, Jonathan E Charnin, and Peter J Pronovost. Identifying meaningful outcome measures for the intensive care unit. *American Journal of Medical Quality*, page 1062860613491823, 2013.

Pamela Mazzocato, Carl Savage, Mats Brommels, Håkan Aronsson, and Johan Thor. Lean thinking in healthcare: a realist review of the literature. *Quality and Safety in Health Care*, 19(5):376–382, 2010.

D Reis Miranda, R Moreno, and G Iapichino. Nine equivalents of nursing manpower use score (nems). *Intensive care medicine*, 23(7):760–765, 1997.

Dinis Reis Miranda, Angelique de Rijk, and Wilmar Schaufeli. Simplified therapeutic intervention scoring system: the tiss-28 items–results from a multicenter study. *Critical care medicine*, 24(1): 64–73, 1996.

San Myint and Mario T. Tabucanon. A multiple-criteria approach to machine selection for flexible manufacturing systems. *International Journal of Production Economics*, 33(1):121–131, 1994.

Nabil Nahas, Mustapha Nourelfath, and Michel Gendreau. Selecting machines and buffers in unreliable assembly/disassembly manufacturing networks. *International Journal of Production Economics*, 154:113–126, 2014.

H. T. Papadopoulos and C. Heavey. Queueing theory in manufacturing systems analysis and design: A classification of models for production and transfer lines. *European Journal of Operational Research*, 92(1):1 – 27, 1996.

H. T. Papadopoulos, Cathal Heavey, and Jim Browne. *Queueing Theory in Manufacturing Systems Analysis and Design.* Chapman & Hall, London, 1993.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

Peter Pronovost, Dale Needham, Sean Berenholtz, David Sinopoli, Haitao Chu, Sara Cosgrove, Bryan Sexton, Robert Hyzy, Robert Welsh, Gary Roth, et al. An intervention to decrease catheter-related bloodstream infections in the icu. *New England Journal of Medicine*, 355(26):2725–2732, 2006.

Peter J Pronovost and George W Bo-Linn. Preventing patient harms through systems of care. *Jama*, 308(8):769–770, 2012.

Peter J Pronovost, Marlene R Miller, Todd Dorman, Sean M Berenholtz, and Haya Rubin. Developing and implementing measures of quality of care in the intensive care unit. *Current opinion in critical care*, 7(4):297–303, 2001.

Peter J Pronovost, David A Thompson, Christine G Holzmueller, Lisa H Lubomski, and Laura L Morlock. Defining and measuring patient safety. *Critical care clinics*, 21(1):1–19, 2005.

Jens Rasmussen and Aage Jensen. Mental procedures in real-life tasks: a case study of electronic trouble shooting. *Ergonomics*, 17(3):293–307, 1974.

James Reason. *Human error.* Cambridge university press, 1990.

HU Rothen, V Küng, DH Ryser, R Zürcher, and B Regli. Validation of "nine equivalents of nursing manpower use score" on an independent data sample. *Intensive care medicine*, 25(6):606–611, 1999.

J. Sadr and R.P. Malhame. Decomposition/aggregation-based dynamic programming optimization of partially homogeneous unreliable transfer lines. *IEEE Transactions on Automatic Control*, 49 (1):68 – 81, jan. 2004. ISSN 0018-9286.

Irvin C. Schick and Stanley B. Gershwin. Modelling and analysis of unreliable transfer lines with finite interstage buffers (Volume VI of Complex materials handling and assembly systems). Report ESL-FR-834-6, MIT Electronic Systems Laboratory, 1978.

J. E. Schor. Efficient algorithms for buffer allocation. MIT EECS MS thesis, Massachusetts Insititute of Technology, May 1995.

Chuan Shi and Stanley B Gershwin. An efficient buffer design algorithm for production line profit maximization. *International Journal of Production Economics*, 122(2):725–740, 2009.

Stevenj Q Simpson, Douglas A Peterson, and Amy R O-Brien-Ladner. Development and implementation of an icu quality improvement checklist. *AACN advanced critical care*, 18(2):183–189, 2007.

A. L. Soyster, J. W. Schmidt, and M. W. Rohrer. Allocation of buffer capacities for a class of fixed cycle production lines. *AIIE Transactions*, 11(2):140–146, 1979.

James A Taylor, Dena Brownstein, Dimitri A Christakis, Susan Blackburn, Thomas P Strandjord, Eileen J Klein, and Jaleh Shafii. Use of incident reports by physicians and nurses to document medical errors in pediatric patients. *Pediatrics*, 114(3):729–735, 2004.

Horst Tempelmeier. Practical considerations in the optimization of flow production systems. *International Journal of Production Research*, 41(1):149 – 170, 2003. ISSN 00207543.

Horst Tempelmeier and Malte Burger. Performance evaluation of unbalanced flow lines with general distributed processing times, failures and imperfect production. *IIE Transactions*, 33(4):293–302, 2001.

J-L Vincent, Rui Moreno, Jukka Takala, Sheila Willatts, Arnaldo De Mendonça, Hajo Bruining, CK Reinhart, PeterM Suter, and LG Thijs. The sofa (sepsis-related organ failure assessment) score to describe organ dysfunction/failure. *Intensive care medicine*, 22(7):707–710, 1996.

Tapio Westerlund and Frank Pettersson. An extended cutting plane method for solving convex MINLP problems. *Computers & Chemical Engineering*, 19:131–136, 1995.