ESSAYS IN HEALTHCARE OPERATIONS

A Dissertation

by

SEUNG JUN LEE

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Gregory R. Heim
Committee Members,	Chelliah Sriskandarajah
	Rogelio Oliva
	James Abbey
	Ramkumar Janakiraman
Head of Department,	Richard Metters

May 2016

Major Subject: Information & Operations Management

ABSTRACT

This dissertation includes three essays, which address significant issues that healthcare practitioners throughout the world face today. The fundamental research that I first address is a research agenda for reimbursement impacts upon healthcare operations management. The purpose of the first essay is to offer conceptual frameworks that portray the fundamental architecture of the U.S. healthcare system and its connections to healthcare reimbursement systems. The research method involves inductive theory development. I contend such frameworks are useful for healthcare operations management research. Using the frameworks, this essay suggests promising research opportunities that should stimulate emerging research themes in the healthcare industry and in academic healthcare operations research. These findings furnish a research agenda with timely insights for practitioners and academia. One conclusion of the essay is the lack of prior research relevant to healthcare reimbursement processes and their impacts on healthcare operations. The essay also concludes that key research opportunities relate to reimbursement boundaries, reimbursement strategy, reimbursement resources, reimbursement impacts, and reimbursement technology.

In the second essay, I examine how scheduling policies can improve healthcare quality and doctor efficiency in outpatient healthcare facilities. The purpose is to develop an outpatient appointment scheduling approach under situations of patient no-shows and patient heterogeneity. Based on detailed analytical and simulation methods, the essay evaluates and compares the performance of my approach against several outpatient scheduling policies under various scenarios, and provides advice regarding optimal policies for outpatient clinics. The findings show that my pro-

posed scheduling algorithms show efficient scheduling performance relative to prior proposed policies. In short, the findings of the second essay provide new applicable scheduling polices for outpatient scheduling. The findings also derive qualitative implications for clinic schedulers for improving the most effective way of scheduling outpatient operations. The conclusion is that the proposed scheduling approach can be potentially useful for outpatient facilities.

Finally, the third essay empirically examines how managerial operational responses of hospitals vary in response to external pressures imposed upon them by government policies. The purpose is to examine whether hospitals respond to such policies by improving operating processes and quality outcomes, or by gaming their response by adjusting patient case mixes and other metrics associated with financial benefits for the hospital, instead of operational improvement. To validate whether hospitals respond suitably to an ongoing U.S. government quality improvement program, called the Value Based Purchasing (VBP) program, I explore how the program influences subsequent behaviors of U.S. hospitals. Using observational data from the Center for Medicare & Medicaid Services (CMS) and several other sources, I use regression analysis methods to provide empirical evidence of the effects of this government policy. The essay findings show that financially penalized hospitals use tactics consistent with symbolic practices, which may be an unintended outcome from the VBP project. The conclusion is that theoretically motivated contextual differences exist in the behaviors of hospitals when facing these external government pressures.

ACKNOWLEDGEMENTS

It would not have been possible to successfully finish my Ph.D. without the deepest help, full support, expert guidance, and encouragement of my advisor Dr. Gregory R. Heim. Without his mentoring in which he always provided a timely and critical manner, my doctoral program would been an immensely frustrating pursuit. Being a great role model in my career, he shared all his wisdom with me and guided me on how to become a good faculty member. Thus, I would like to express my deepest gratitude to Dr. Gregory R. Heim. I cannot imagine a better advisor than Dr. Heim.

I would like to express my sincere appreciation to my dissertation committee members who provided endless support and held me to a high standard. I am ineffably indebted to Dr. Chelliah Sriskandarajah for conscientious guidance and encouragement that enabled me to reach beyond the boundaries of empirical methods. Without his patience and warm encouragement, I would not have successfully completed my scheduling project. My successful research progress also owes a lot to Dr. James D. Abbey and his father, Dr. Duane Abbey, who shared practitioner-oriented knowledge relevant to my first essay. Without their deep knowledge in healthcare reimbursement, I would not have completed my first essay. I am also grateful for James' constant support and encouragement outside of the research context.

I am most grateful to Dr. Rogelio Oliva for encouraging me to deepen my research skills. His valuable guidance and support enabled me to complete my dissertation. In addition, I thank Dr. Ramkumar Janakiraman who suggested I consider various modeling approaches in my empirical research. Also, I would like to express special thanks to Dr. Xenophon Koufteros who provided great advice and encouraged me to seek managerial insights during my doctoral program.

Above all, I thank my wife, Jeong Hyun Park for her personal support and great patience at all times. My proud son David and lovely daughter Claire have been giving me an incredible motivation all the time. My parents, parents-in-law, and brother have given me their unequivocal support. My mere expression of thanks does not suffice to thank all my family members.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

1. INTRODUCTION

The large and growing body of research making up extant healthcare service research has a general consensus that current healthcare systems suffer from many inefficiencies of healthcare delivery (Green, 2012). In response, governmental bodies and healthcare institutions are trying to ameliorate the inefficiencies of the healthcare system through a mix of new processes, policies, incentives, and penalties. This dissertation examines several healthcare operations management topics inspired by these developments. The dissertation comprises three essays concerning conceptual foundations for healthcare operations management research, severe healthcare service scheduling problems, and healthcare organizational responses to incentive policies. Specifically, the dissertation develops conceptual frameworks of the healthcare system and healthcare reimbursement processes, analytical models of outpatient scheduling processes, and econometric models for hospital procurement behaviors in response to government financial incentives.

A fundamental yet untouched area of research that I first address concerns a research agenda for healthcare operations management motivated by the healthcare reimbursement process. The first essay examines reimbursement processes within the U.S. healthcare system to motivate research opportunities for operations and supply chain management (OM/SCM) researchers. Healthcare reimbursement processes consist of coding, billing, and payment processes related to care provided. These processes have significant implications for nearly three trillion dollars of annual financial flows, making up the largest single sector of U.S. gross domestic product (GDP). To begin to address the implications of such enormous financial flows for healthcare operations managers, this essay uses conceptual frameworks to illustrate the complexity of multi-stakeholder healthcare reimbursement processes, identify salient operating challenges, and develop an agenda of research opportunities. Healthcare organizations face numerous operating challenges and decreasing reimbursement rates due to many regulatory and market pressures. Healthcare providers must conform to regulations and policies of multiple external organizations that have power to exact control over providers through both service provision and reimbursement. The often myopic focus in extant research represents a significant academic gap in understanding the broad nature of reimbursement processes, regulations, and organizations. Healthcare service providers must adopt and conform their operations processes to many reimbursement systems. The multiplicity of healthcare providers and reimbursement systems generates operational complexity and uncertainty, which this essay illustrates with end-to-end conceptual frameworks of healthcare reimbursement processes.

In the second essay, I consider an outpatient appointment scheduling system involving patient no-shows and patient heterogeneity to tackle healthcare process scheduling issues that lead to inefficient performance and financial outcomes. In the current outpatient scheduling systems, patients are suffering from long waiting times while physicians are complaining about their overwork and overtime hours (Cayirli et al., 2006). Accordingly, these factors may result in healthcare delivery system operational failures, leading to worse healthcare outcomes. For example, Veterans Affairs (VA) administrators falsified VA patient scheduling data, and reported that VA facilities had satisfied required performance standards (Kensling and Nissenbaum, 2014). The root cause of the VA facilities scandal is an unbalance between patients' prolonged waiting times and the available healthcare capacity. This practical issue and a gap within appointment scheduling research triggers the need for exploring this particular project.

Specifically, I study block scheduling policies for single providers under conditions of patient heterogeneity in service times and patient no-shows. The research objective is to find daily appointment schedules that minimize a weighted sum of patients' waiting time, the physician's idle time, and the physician's overtime. Compared to extant outpatient scheduling approaches, this essay contributes by suggesting new sequential block scheduling procedures grounded in actual outpatient clinic practices and in the successful Toyota Production System load smoothing approach, leading to effective appointment schedules when scheduling two heterogeneous patient types. The proposed block scheduling policy first assigns a sequence of different patient types, given patient demand and service time information. The policy then allocates repetitive time blocks in a planning horizon. Using the block scheduling policy, I examine different scenarios that outpatient clinics face, including patient overbooking and open-access scheduling policies. The proposed approach is found to generally perform better than methods proposed in prior work.

Finally, the third essay examines healthcare financial incentive and penalty policy to determine whether one such program has changed practices and processes of healthcare providers in the intended manner. Healthcare organizations still face process and care outcome inefficiencies. With the Affordable Care Act (ACA) launch in 2010, U.S. healthcare policy required hospitals to focus on patient safety, care quality, and process improvement. The Value Based Purchasing (VBP) program, one of several federal regulations directed by Medicare based on a financial components approach for healthcare reimbursement, encourages hospital managers to enhance process quality, improve patient satisfaction, and improve care outcomes. Prior to the VBP program implementation, many practitioners claimed VBP would have little impact or would lead to the unintended consequence of harming previously poor-performing hospitals by instead giving financial incentives to well-off hospitals. Despite the expanding academic study of healthcare operations management, little research clarifies the practical behaviors of care provider operations when they modify practices and processes in response to external financial regulatory pressures. By combining secondary data sets from several sources, I empirically examine impacts of VBP penalties on subsequent hospital behaviors. The study finds that financially penalized hospitals are more likely to adopt symbolic management practices, as represented by changes in their patient case mix and two additional monetary incentive related measures.

My dissertation has several contributions in the operations and supply chain management area. In the first essay, the study highlights unexplored operational process areas that have yet to be examined by healthcare OM/SCM scholars. Based on the agenda of research opportunities the essay presents, operations management researchers in the healthcare domain may need to extend their managerial interest to healthcare reimbursement processes. Scholars cannot account for all of the factors that enable or hinder quality healthcare service design and delivery if researchers ignore or improperly account for reimbursement processes. I believe that the identified research opportunities will gain increased attention and become an important problem domain in the near future. Specifically, healthcare service OM/SCM researchers have rarely, if at all, examined healthcare reimbursement processes and related operational issues. Thus, the essay provides directions for future studies into financial process impacts upon service operations within the healthcare industry. In addition, the essay aims to suggest high-level guidance for practitioners. In particular, my research frameworks may allow healthcare managers to identify and conceptualize the antecedents and consequences of administrative reimbursement issues, which could result in reduction of overhead and medical error costs. As the U.S. healthcare system is among the most complicated in the world (Klepper, 2011), many components of the frameworks may prove useful for understanding other countries' simpler but related healthcare reimbursement systems. If academic researchers provide constructive research questions and corresponding suggestions, the healthcare field might be further improved in terms of efficient and effective operations management. I believe that my frameworks related to sources of complexity and uncertainty, and corresponding research opportunities, will contribute important research questions and unexplored issues to the OM/SCM field.

The second essay contributes by introducing a simple and easy-to-use block scheduling policy that repeats block assignments throughout a day. This block scheduling policy is based on actual practices of outpatient schedulers as well as in the very successful manufacturing approach used in the Toyota Production System for scheduling multiple product types. In particular, patient heterogeneity and patient no-shows are two factors that practitioners in outpatient services are interested in (Huang and Verduzco, 2015). In the U.S., for example, there are over 900 million outpatient ambulatory care visits annually (CDC and Prevention, 2010), at over 96,000 outpatient care centers establishments (Business Data Codes, 2015). Thus, this essay conveys managerial insights that may prove useful to many outpatient clinic managers and schedulers. Finally, this essay can also be widely applicable to other professional service organizations (e.g., financial consultations) in the context of scheduling customers of multiple types having relatively fixed service times.

Finally, the third essay contributes by providing theoretical arguments to explain operational behaviors of hospitals when facing external government pressures such as VBP. As traditional supply chain management studies have investigated interconnected supply chain activities to improve the value of a supply chain, this essay also will contribute by exploring incentive alignment and coordination problems (i.e., between hospitals and third party payers) within the healthcare supply chain. By empirically examining the evidence of symbolic practice and drivers of this practice, this essay extends the institutional and symbolic management perspective into healthcare service operations management research. Next, there is little empirical research in healthcare operations management that examines responses to various institutional pressures, thus this essay contributes to empirical evidence by quantifying the impact of VBP. As Green (2012) suggests, managing patient-oriented service processes is an essential topic for the future of the operations management field. Since the VBP program is a touchstone program intended to accomplish patient-oriented service delivery and to improve healthcare outcomes, my empirical analysis of the VBP program contributes to this aim.

In summary, my dissertation develops theoretical contributions in healthcare operations supported by grounded theory from practitioners' insights and extant literature, such as organizational theory, economic theory, and scheduling theory. Using three individual essays in the healthcare operations management domain, I create a research portfolio that considers theoretical contributions, empirical analyses, and analytical scheduling contributions. Based upon managerial problem-motivated research, the following three chapters articulate timely and critical issues for both practitioners and academic researchers in the healthcare industry.

The remainder of this dissertation is structured as follows. Section 2 provides conceptual frameworks of the healthcare system and healthcare reimbursement processes. Section 3 develops block scheduling models for outpatient scheduling processes. Section 4 explores econometric models for hospital procurement behaviors in response to government financial penalties. Section 5 briefly concludes this dissertation.

2. HEALTHCARE OPERATIONS AND REIMBURSEMENT SYSTEM PROCESSES: A RESEARCH AGENDA FOR OPERATIONS AND SUPPLY CHAIN MANAGEMENT

2.1 Introduction

This chapter develops conceptual frameworks of reimbursement processes within the U.S. healthcare system to motivate a research agenda for operations and supply chain management (OM/SCM) researchers. Healthcare reimbursement processes include coding processes, billing processes, and payment processes. Our extensive review of literature on healthcare services reveals that few OM/SCM researchers examine reimbursement processes. Yet, scholars and healthcare practitioners hold a consensus that existing U.S. healthcare reimbursement systems have a complex and awkward structure (David, 2014; Institute of Medicine, 2001; Rouse and Serban, 2014), which triggered a 10.1% rate of incorrect Medicare reimbursements accounting for $36 billion in incorrect payments in 2013 (Adamy, 2014). Clearly, improving healthcare services and ensuring correct and timely provider compensation requires understanding reimbursement processes, associated errors, and consequent care delivery implications.

Avoiding harmful reimbursement system consequences is societally important for parties involved in healthcare consumption, provision, and financial flows. Healthcare providers face increasing process variation, which can decrease service quality and increase healthcare costs (Schmenner, 2004; Tucker et al., 2007). U.S. healthcare expenditures grew from $2.2 trillion in 2007 (Barton, 2010) to nearly $3 trillion in 2013, now the single largest sector of U.S. Gross Domestic Product (GDP) at 17.9% (WorldBank, 2014). As many argue, continued growth of healthcare costs will harm various business sectors and the economy as a whole (Baker and Rosnick, 2005; Brill, 2015).

Medical errors represent one source of significant, wasteful increases in healthcare costs. Healthcare providers face two types of medical errors. One error type relates to clinical errors, such as diagnostic or surgery mistakes. Clinical errors present significant operations problems affecting care delivery. In the 1990s, the annual cost of U.S. clinical errors was $23 billion, and more than 7,000 patients died from the errors (Kohn et al., 1999). Clinical errors today still cause major injuries or death in up to 160,000 people yearly (Landro, 2013) and have motivated much OM research.

The other error type relates to administrative errors, which include incorrect coding, unsuitable billing, or inadequate payment documentation by healthcare providers. Administrative billing and claims filing errors present significant operations problems that directly and indirectly affect care quality, yet few OM/SCM studies examine these reimbursement issues. Recent evidence suggests about 30% of medical billing claims contain errors (Silver-Greenberg, 2011). Healthcare administrators and government policy makers spotlight reimbursement process errors as serious causes of poor operational performance (O'Malley et al., 2005). In light of these issues, advancing literature on healthcare reimbursement processes is of rising societal and economic importance. Green (2012) identified the need for evidence–based healthcare research using operational research methods, which will require a new focus on healthcare reimbursement and the operational hazards of administrative errors. A necessary first step for researchers involves understanding the scope and complexity of reimbursement processes.

Compared to prior literature, this chapter contributes and provides guidance for future research by considering various reimbursement processes and related operations through conceptual frameworks that detail the overall flow of services and financial transactions within healthcare. This study also provides high–level guidance for reimbursement process managers. Specifically, the conceptual frameworks may enable physicians, managers, and chief compliance officers (CCO) to better understand operational drivers and consequences of reimbursement process problems, which may help reduce overhead and costs.

The following sections articulate foundational literature, conceptual frameworks, and research opportunities in healthcare reimbursement systems. For readers unfamiliar with the broad range of reimbursement terminology, Table A.1 in Appendix A provides definitions for various acronyms and terms.

2.2 Background and Related Literature

This section first reviews healthcare literature predominantly related to three topics: a brief overview of healthcare reimbursement literature followed by more in–depth examination of healthcare service and healthcare information technology (IT) literature. As this section demonstrates, a void in research regarding healthcare reimbursement processes offers significant opportunities for academic studies to bridge the identified research gaps.

2.2.1 Healthcare Service Literature

Even though there are several different types of healthcare reimbursement systems, much of the extant literature typically assumes the use of one emblematic system, the Diagnosis—Related Group (DRG) system (Green, 2012; Powell et al., 2012) and often does so only tangentially in relation to non–reimbursement research questions. The DRG system is one payment mechanism (i.e., a prospective payment system (PPS)) established by the Centers for Medicare and Medicaid Services (CMS)—a federal agency that manages the Medicare and Medicaid programs in cooperation with state governments. Under a PPS, third party payers (TPPs) reimburse providers via a fixed schedule of payment rates for individual services provided (Roth and Van Dierdonck, 1995). The DRG system only covers a limited set of inpatient healthcare services. CMS operates other PPSs for hospital outpatient services, such as Ambulatory Payment Classifications (APCs), Skilled Nursing Facilities (SNFs), Home Health Agencies (HHAs), and Long—Term Care Hospitals (LTCHs) (Abbey, 2012). Outpatient expenses have grown by 10.1% annually—double the current growth rate of inpatient DRG expenses—emphasizing the need for researchers to account for impacts of these multiple reimbursement systems (HCCI, 2010). Accumulating evidence also suggests various adverse impacts of complex reimbursement processes on patient care provision (AAPS, 2000; Taylor and Morrison, 2011) and on decreasing doctor career satisfaction and retention (Dougherty, 2001). Interestingly, operations management, supply chain, and information systems (IS) research seldom tackles such issues. As the U.S. healthcare system is one of the most complicated (Klepper, 2011) and the system adopts standardized worldwide treatment codes (e.g., ICD–10), such research may prove useful in understanding reimbursement systems employed by other countries (Quan et al., 2008).

Conceptual OM/SCM healthcare service literature defines detailed lists of medical issues at strategic, tactical, and operational levels (Hans et al., 2012) and describes healthcare operations characteristics, such as patient types, service types, and performance measures (Cayirli et al., 2006; Gupta and Denton, 2008; May et al., 2011). Motivated by prior literature (AMA, 2013; Hulshof et al., 2012), our review focuses on five classes of healthcare service processes: emergency care, inpatient care, outpatient ambulatory care, home and residential care, and administrative services. As systems approaches facilitate quality improvement (Chandrasekaran et al., 2012; Flynn et al., 1994), we also carefully reviewed the three process tiers within healthcare delivery and reimbursement systems: patients, care providers, and TPPs, to identify extant studies of impacts of reimbursement processes on the flow of healthcare. To expand upon prior reviews focused on specific services (e.g., emergency), or methodologies (e.g., scheduling), we reviewed conceptual, analytical, and empirical studies, and studies on all healthcare service classes. Table 2.1 contains literature related to within hospital/clinic (i.e., intra–firm) service processes. Table 2.2 shows literature related to healthcare supply chain (i.e., inter–firm) processes. Each table identifies service class, dominant concepts, and whether the study considered reimbursement processes.

Table 2.1: Related literature on intra–firm (within hospital/clinic systems) healthcare research

Literature Category	Authors	Healthcare Classes	Dominant Concepts	Reimbursement System
Focus on Physicians' Treatment Process Efficiency and Effectiveness	Clark and Huckman (2012)	Inpatient	Focused operations related to co-specialization in related areas provide positive quality performance, but no empirical evidence due to specialization in specific practice.	None
	Dobson et al. (2009)	Inpatient	Physicians are less likely to gain financial benefit from assigning work to frontline staff.	None
	Jiang et al. (2012)	Ambulatory	Performance-based system is superior to FFS contracts.	U.K. System
	KC (2013)	Emergency	Physician multitasking influences processing time in a U-shaped manner.	None
	KC and Terwiesch (2011)	Inpatient	Hospital focus is associated with clinical performance, such as outcomes and quality.	None
	KC and Terwiesch (2012)	Inpatient	Patient early discharge is associated with readmission, and when occupancy level is high in the hospital, patients are more likely to be discharged.	None
	Nair et al. (2013)	Inpatient	Clinical quality and flexibility improve operational efficiency while experiential quality moderates the association.	DRG

Table 2.1 Continued

Literature Category	Authors	Healthcare Classes	Dominant Concepts	Reimbursement System
	Powell et al. (2012)	Inpatient	When doctors spend more time on paper work, medical billing errors will decrease.	DRG
	Song et al. (2015)	Emergency	Physician use of dedicated queuing can enhance physician ownership over patients and improve operational efficiency.	None
Focus on Nurses' and Other Factors' Efficiency and Effectiveness	Angst et al. (2011)	Administrative	The sequence of healthcare IT implementation influences hospital performance effects.	None
	Boyer et al. (2012)	Inpatient	In small hospitals, focusing on specific outcomes related practices provides better quality, while larger hospitals work better with climate focused on specific outcome goals.	None
	Chandrasekaran et al. (2012)	Inpatient	Hospital process management is negatively associated with patient satisfaction (experiential quality), but patient-centered leadership can mitigate this negative relationship.	None
	Ding (2014)	Administrative	U.S. hospitals follow organizational learning curves for productive efficiency. Operational focus in a hospital can lead to productive efficiency	DRG
	Goldstein and Iossifova (2012)	Inpatient	Organizational slack consisting of available and accessible resources in a hospital affects process performance.	None
	He et al. (2012)	Inpatient	Reducing hospital labor costs through the timing of staffing decisions under uncertainty about nurses' daily workload.	None
	Jack and Powers (2004)	Administrative	Volume flexibility can be a strategic choice to tackle demand uncertainty in healthcare.	DRG
	Lahiri and Seidmann (2012)	Administrative	When a hospital fails to obtain necessary clinical information, this causes a significant impact on the total turnaround time of diagnostic reports.	None
	Marley et al. (2004)	Administrative	Leadership is positively associated with clinical and process quality, and hence affects patient satisfaction.	None

Table 2.1 Continued

Literature Category	Authors	Healthcare Classes	Dominant Concepts	Reimbursement System
	Shortell et al. (1995)	Inpatient	Find little effect of TQM and organizational culture on cardiovascular patients	None
	Silverman and Skinner (2004)	Administrative	Care providers are likely to use more upcoding to increase profits.	DRG
	Spear (2005)	Administrative	Learning to improve process quality while professionals actually deliver the service can reduce medical errors.	None
	Tucker (2007)	Inpatient	Nurses' team-based initiatives can improve the work system.	None
	Tucker et al. (2007)	Inpatient	Hospital teams that focus on learn-how (activities that operationalize practice in a given setting) may achieve more implementation successes than teams who focus on learn-what (activities that identify current best practices).	None

Table 2.2: Related literature on inter-firm healthcare supply chain research

Literature Category	Authors	Healthcare Classes	Dominant Concepts	Reimbursement System
Focus on Healthcare Supply Chain Management	Angst et al. (2010)	Administrative	Diffusion of electric medical records in hospitals is associated with susceptibility to the influence of prior IT adopters.	None
	Chen et al. (2013)	Administrative	Healthcare supply chain performance is associated with supplier integration, which consists of knowledge exchange and IT integration.	None
	McKone-Sweet et al. (2005)	Administrative	Lack of executive support and misaligned incentives between chain members can be barriers for improving healthcare supply chain	None
	Sinha and Kohnke (2009)	Administrative	Affordability, access, and awareness framework for the healthcare supply chain.	General Finance

Table 2.2 Continued

Literature Category	Authors	Healthcare Classes	Dominant Concepts	Reimbursement System
Focus on Environmental and Policy Related Literature	Bhakoo and Choi (2013)	Administrative	Hospital symbolic practice varies depending on the characteristics of institutional pressures	None
	Fuloria and Zenios (2001)	Administrative	The outcomes-adjusted reimbursement system, which is based on adverse short-term patient outcomes can maximize social welfare.	Fee for service
	Lee and Zenios (2012)	Administrative	The standalone operating incentive payment system is not comprehensive to improve healthcare delivery system.	Operating Incentive Payment System
	Meyer and Collier (2001)	Administrative	Estimates causal relationships in the quality management program (Malcolm Baldrige National Quality Award Criteria).	None
	Miller and Tucker (2009)	Administrative	State privacy regulations reduce the diffusion of electronic medical records (EMRs) in hospitals.	Fee for service

Our review of healthcare service literature suggests various prominent characteristics. One characteristic is the dominance of studies about particular healthcare service classes (i.e., inpatient services), with a lack of findings pertaining to other healthcare service classes. Another characteristic relates to lack of literature focusing on financial flows within healthcare reimbursement systems and their impact on patient care delivery (Roth and Van Dierdonck, 1995). Building on the limited literature base, four primary research gaps provide actionable impetus for exploring reimbursement process issues. First, healthcare service research mostly focuses on care delivery processes and corresponding care outcomes. Few studies (Sinha and Kohnke, 2009; Powell et al., 2012) lay out boundaries of healthcare reimbursement issues, revealing a lack of clear contexts for precise research and construct definition.

Second, though the field of medicine has benefited from the integration of theory development and empirical research (Fisher, 2007), there appears to be a lack of theory pertaining to the business operations portion of medicine, that is, healthcare reimbursement processes and outcomes. Third, although some seminal studies examine salient resources within care delivery processes (Tucker, 2007; Tucker et al., 2007), there is a broad research gap regarding resources (e.g., accountable care organizations, code professionals, RACs, IT) for reimbursement processes. Fourth, few analytical or empirical studies investigate impacts of healthcare reimbursement processes (Powell et al., 2012).

Next, we review related healthcare IT (HIT) research (Table 2.3), as IT often is assumed to have resolved healthcare reimbursement issues. Most healthcare systems and care providers have adopted internal IT applications, such as electronic health record (EHR) software systems for collection of electronic patient health information, to improve patient care quality (Bardhan et al., 2014). Without IT, modern healthcare services often cannot operate smoothly (Angst and Agarwal, 2009). Many healthcare IT researchers investigate antecedents and roles of IT (see Table 2.3). Such studies often focus on IT adoption and diffusion effects (Angst et al., 2010) or the impact of IT usage on patient care delivery (Devaraj and Kohli, 2003). The studies execute hospital–level (Agarwal et al., 2010; Das et al., 2011), or individual physician–level research (Angst et al., 2010; Gao et al., 2012). The findings often show healthcare IT adoption is positively associated with healthcare quality and efficiency (Buntin et al., 2011).

Table 2.3: Related literature on IT healthcare research

Literature Category	Authors	Healthcare Classes	Dominant Concepts	Reimbursement System
Focus on Healthcare IT Adoption	Agarwal et al. (2010)	Administrative	Identify three major areas for future research in HIT.	None
	Angst and Agarwal (2009)	Administrative	The likelihood of care providers electronic health records (EHR) adoption is associated with individuals' concern for their information privacy.	None
	Angst et al. (2010)	Administrative	Diffusion of electric medical records in hospitals is associated with susceptibility to the influence of prior IT adopters.	None
	Angst et al. (2011)	Administrative	The sequence of healthcare IT implementation influences hospital performance effects.	None
	Goh et al. (2011)	Administrative	To implement HIT successfully, hospitals should manage co-evolution process between work routines and HIT.	None
	Kallinikos and Tempini (2014)	Administrative	The web-based patient self-reporting data can become new models of organizing medical knowledge creation.	None
	Yaraghi et al. (2014)	Administrative	Care providers' HIE adoption behaviors are related to topographical factors, isomorphic effects between providers, and labor inputs in HIE use.	None
	Kohli et al. (2012)	Administrative	IT investment positively affects a hospital's market value, such as financial and accounting performance measures.	None
Focus on the Impact of Healthcare IT	Bardhan and Thouin (2013)	Administrative	Clinical information systems and patient scheduling applications are positively associated with three major clinical outcome metrics, such as heart attacks, heart failures, and pneumonia.	None
	Bardhan et al. (2014)	Inpatient	Health IT usages are associated with reducing patient readmission risk.	None
	Buntin et al. (2011)	Administrative	Many care providers (more than 90%) have positive impacts of HIT.	None
	Devaraj and Kohli (2003)	Administrative	IT usage is positively associated with hospital performance, such as revenue and quality.	None

Table 2.3 Continued

Literature Category	Authors	Healthcare Classes	Dominant Concepts	Reimbursement System
	Devaraj et al. (2013)	Administrative	IT is positively associated with patient flows. In particular, swift patient flow can affect financial performance while even patient flow can affect quality performance.	None
	Gao et al. (2012)	Administrative	Identified that there is no evidence that online physician rating systems are dominated by dissatisfied patients.	None
	Kohli and Kettinger (2004)	Administrative	With IT adoption, care providers can reduce clinic procedural costs and enhance transparency.	None
	Mukhopadhyay et al. (2011)	Administrative	Learning rates of IT-enabled physician referral systems differ across different agents with different skills, and IT system upgrade has a positive impact on the performance of experts.	None

The review of extant IS research reveals other research gaps. Although many studies examine adoption and impacts of healthcare IT applications (Agarwal et al., 2010), little literature considers IT applications pertaining to reimbursement processes. Indeed, academic studies rarely consider non–care–related IT modules, such as physician referral systems (Mukhopadhyay et al., 2011). Specifically, few IT studies focus on supply chain-level healthcare reimbursement flows and corresponding care impacts. For instance, to date, no healthcare IT literature has examined moderating or mediating effects of reimbursement systems.

Our review of the literature suggests two significant gaps. First, certain healthcare service classes (i.e., inpatient or administrative services) dominate the OM/SCM healthcare literature. Second, little literature focuses on financial flows moving through healthcare reimbursement systems. These findings suggest a need for research pertaining to other healthcare service classes and related reimbursement issues. Subsequent sections address this research need by illustrating the scope and complexity of healthcare reimbursement processes.

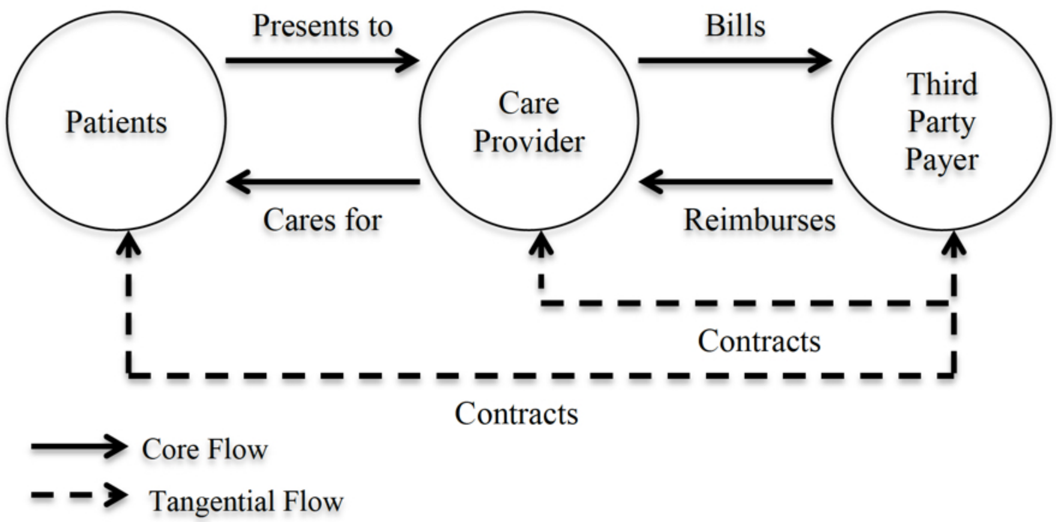
2.3 Conceptual Frameworks of Healthcare Delivery and Reimbursement Processes

We start with simple models of the overall healthcare system, and move step–by–step toward increasingly complex models of reimbursement processes. In doing so, this section provides conceptual foundations upon which to build constructs and research opportunities.

2.3.1 Process Model of Care Delivery and Reimbursement Processes

We begin with a simple model of the overall healthcare system in terms of flows of care and resulting reimbursement processes. Figure 2.1 deconstructs healthcare operations into core and tangential service flows. The first key process is the core flow (solid lines) between patients and care provider—the left two parties of Figure 2.1. When an individual seeks a healthcare service, the individual selects a care provider, such as a hospital or clinic. After an individual presents to a care provider, and subsequently checks in, the individual is defined as a patient. At this point, an encounter begins. For purposes of this essay, an encounter refers to direct contact between the patient and an authorized care provider (e.g., a physician) for the purposes of diagnosis or treatment of the patient (Medicare Claims Processing Manual, 2008). In general, there must be face–to–face contact between the authorized care provider and patient to bill as an encounter, though some TPPs may have different requirements. In most cases, when a patient requires more than one follow–up visit with a care provider, each subsequent visit is considered a separate encounter for billing purposes. During the encounter, the care provider provides diagnostic or treatment services to the patient. The patient checks out and leaves the provider

Figure 2.1: Healthcare service delivery system flows



after completion of the service provision encounter. The second core flow occurs between a care provider and TPP. When services are provided, the care provider codes the performed care procedures into its healthcare record system, develops a bill, translates the bill into a claim, and the claim is filed to a TPP. Finally, the service can be completed after the TPP reimburses the provider for the claim. An individual who received treatment may pay bills directly to a provider. For self–pay patients, the diagram may not require the TPP, as the billing process occurs between a patient and a care provider, except in cases where the patient submits a bill directly to the TPP for reimbursement. Even in cases where a self–pay patient cannot pay for a service, a care provider may still need to perform a required limited service under healthcare law (e.g., EMTALA, see Table A.1 in Appendix A).

The two core flows are supported by two tangential service flows (dashed lines). Both tangential flows are associated with a TPP. One tangential flow represents the contract between patients and a TPP. Individuals choose and enroll with one or more

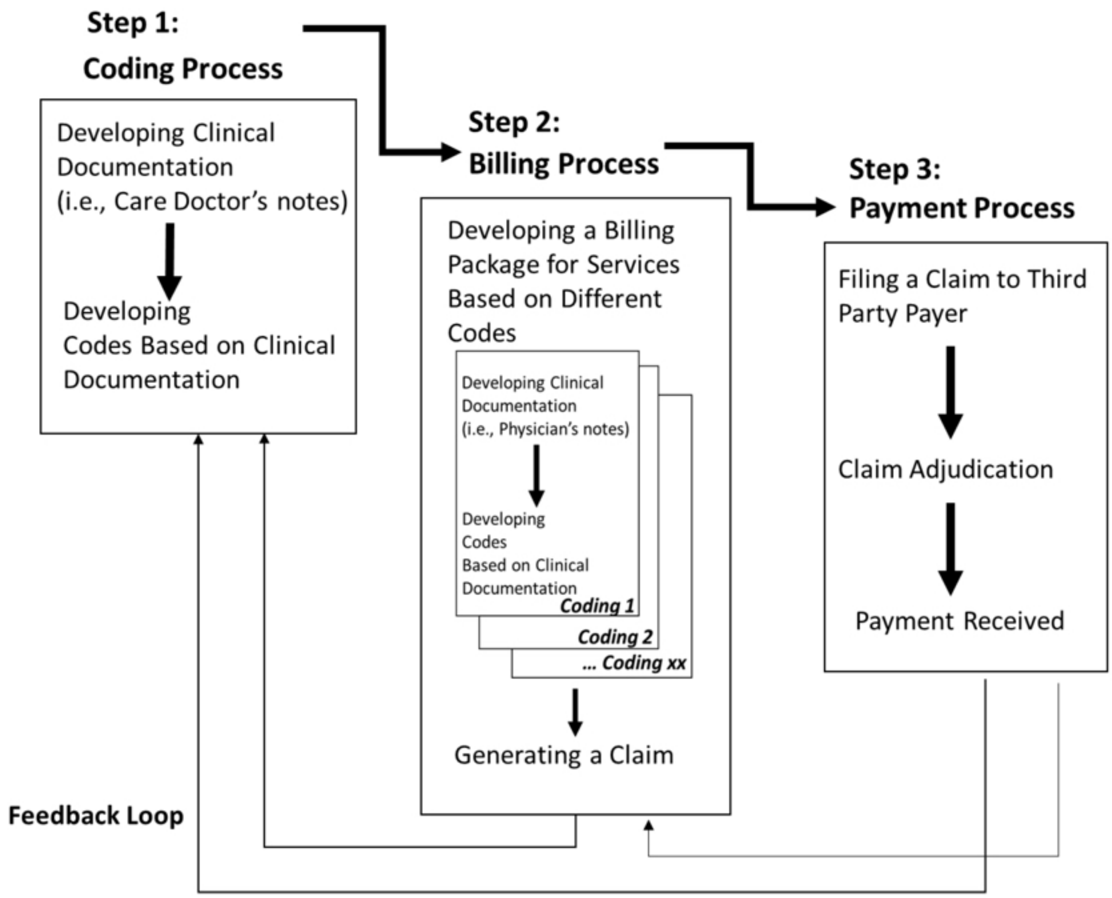
TPPs before receiving services. Typically, an individual makes regular payments to the TPP. When individuals become patients, the TPP pays for the covered services as defined in the contract. The other tangential flow is a contract between a care provider and a TPP. From the patient perspective, this tangential flow constrains the patient's choice of a care provider because, in many cases, patients only visit a preferred care provider that has a contract with their TPP(s).

A combination of core and tangential flows describes the healthcare system service delivery and financial flow mechanisms. Prior healthcare OM/SCM research explores quality management and process improvement within the leftmost core flows (Boyer et al., 2012; Dobrzykowski et al., 2014). However, no extant literature examines the full set of healthcare system flows, including reimbursement. We next delve into the reimbursement process structure.

2.3.2 Process Model of Reimbursement Processes

As defined in Section 2.1, reimbursement processes involve three steps—coding, billing, and payment—starting with the development of clinical documentation by physicians who participate in care provision. Figure 2.2 diagrams the care provider's reimbursement process. In the first step–the coding process—a physician (or another licensed care provider) files a patient's medical record for the provided services and items supplied for the patient's treatment. The physician's documentation then becomes available for professional coding staff in the reimbursement (or billing) department of a care provider. Coding staff then develops a set of appropriate codes—alphanumeric string—representing services and facilitating material inputs representing the care provided. The coding process can be quite complex due to complicated coding guidelines. The second step is the billing process. Since one service may require two or more codes, the professional coder develops a billing package

Figure 2.2: Care provider reimbursement process



for the portfolio of services provided. The coder then prepares a claim to collect payment for the billing package. Finally, in the third step, the payment process entails filing a claim to a TPP, claim adjudication, and eventually—in theory—final payment.

2.3.3 Types of Healthcare Organization Entities

Table 2.4 delineates subcategories of patients, care providers, and TPPs as a means to further characterize entities involved in reimbursement in Figures 2.1 and 2.2. For communicating and describing healthcare services, patients divide into

Table 2.4: Types of organizational entities

Type of Patients	Types of Care Providers	Types of Third Party Payers
Office Visits	Inpatient Hospitals	Governmental Organization
New Patient	Outpatient Hospitals	Medicare
Established Patient	Ambulatory Surgical Centers	Medicaid
	Clinics	Veteran Affairs
Types of Service	Home Health Agencies	Indian Health
Inpatient	Durable Medical Equipment Suppliers	Community Health
Outpatient	Other Facilities	Migrant Health
Resident		
		Private Insurance Organizations
		Employment-Related Insurance
		Other Private Insurance

one of two types: new patients or established patients (AMA, 2013). According to care provider guidelines, a new patient is defined as a recipient who has not received services from the same care provider within a certain time period. Otherwise, the patient is an established patient. Patients are also subdivided by their length of stay in a care provider (e.g., hospital, clinic, or associated facility). A patient who stays more than 24 hours is classified as an inpatient whereas a patient who visits less than 24 hours is an outpatient. A patient may receive services in the patient's home, such as in the case of home health agencies, or a patient may become a resident at a nursing home in order to receive care. For extended care service, physicians generally develop and certify a plan of care (POC) to establish the boundaries, timeline, and requirements for continued care (Medicare, 2015).

The second column of Table 2.4 outlines the breadth of healthcare providers. Criteria of Medicare specify care provider classes—inpatient hospitals, outpatient hospitals, ambulatory surgical centers, clinics, home health services, durable medical equipment suppliers, and other facilities—based on care provider capabilities (eCFR, 2015). Hospitals often provide both inpatient and outpatient services. Ambulatory surgical centers (ASCs) are independent facilities that treat patients who need ambulatory surgeries taking less than 24 hours (ASCA, 2013). Clinics are care providers similar to outpatient hospitals, but the organizational entity in terms of payment can be physician groups (Abbey, 2009). Home health agencies (HHAs) are care providers for patients who receive care services in their homes. Durable medical equipment (DME) suppliers provide medical products that a doctor prescribes for patient care, including devices such as arm braces, blood glucose monitors, or wheelchairs (Medicare, 2014). Though additional types of healthcare service providers exist (Abbey, 2009), the five types in Table 2.4 provide a baseline to understand reimbursement processes.

The third column of entities concerns TPP types. The term TPP represents an entity involved in healthcare payment, where the entity is neither a patient nor a care provider (Health Law Resources, 2012). For instance, a TPP is any individual, entity, or program that is, or may be, liable to pay for any medical assistance provided to a recipient (OIG, 2014). For simplicity, Table 2.4 ignores liable individuals. TPPs, largely comprised of various insurance firms, divide into two general categories: government organizations (e.g., Medicare, Medicaid) and private insurance organizations (e.g., Blue Cross Blue Shield). As Table 2.4 shows, the range of TPPs is quite broad.

2.4 Breadth of Healthcare Payment Systems

Though previous literature examines specific payment systems (Powell et al., 2012) or payment policies (Lee and Zenios, 2012), no extant healthcare OM literature presents a comprehensive framework of types of payment systems within reimbursement processes. Thus, Table 2.5 identifies a broad set of U.S. healthcare payment systems used by TPPs to reimburse providers. Most healthcare payment systems in the U.S. are based on fee–for–service, wherein services are reimbursed based upon the quantity of care, and a TPP pays for the healthcare services (Abbey, 2013; Berenson and Rich, 2010). As a baseline, fee–for–service systems can be subdivided into four categories: cost–based payment systems, charge–based payment systems, fee schedule payment systems, and prospective payment systems (PPSs). In addition, many other systems exist, such as capitated, hybrid, managed care, and operating incentive systems. Each system aims to support specific healthcare services and contains tremendous variations depending on its care providers and patient beneficiaries.

Table 2.5: Types of healthcare payment systems and specific programs

Payment Type	**Adopts Coding System**	**Coding System Types**	**Applicable Care Providers**
Cost-Based Payment	No	N/A	Pharmaceutical Items, Critical Access Hospitals (CAHs), Rural Health Clinics (RHCs),Federally Qualified Healthcare Centers (FQHCs)
Charge-Based Payment	No	N/A	Limited use by contract. Charge structures also used for PPSs
Fee Schedule Payment			
Resource Based Relative Value System (RBRVS)	Yes	CPT, (or Current Procedural Terminology):Published by the American Medical Association HCPCS, (or Healthcare Common Procedure Coding System): Published by the Centers for Medicare and Medicaid Services	

Table 2.5 Continued

Payment Type	Adopts Coding System	Coding System Types	Applicable Care Providers
Prospective Payment System			
Diagnosis-Related Groups (DRGs)	Yes	DRG categories: Published by the Medicare Program, ICD-9-CM	Hospital inpatient services
Ambulatory Payment Classifications (APCs)	Yes	APCs categories: Published by the Medicare Program, ICD-9-CM	Hospital outpatient services
Home Health and Skilled Nursing Facilities	Yes	Resource Utilization Groups, Published by the Medicare Program, ICD-9-CM	Home health and skilled nursing services
Capitated Payment (Bundled Payment System)	Yes	Both Fee Schedule Payment and PPS coding systems	Some specific services that TPP allows. And limited care providers
Hybrid Payment System	Yes	Both Fee Schedule Payment and PPS coding systems (If needed)	All care providers
Managed Care Payment System	Yes	APC categories and CPT code systems	All care providers
Operating Incentive Programs		N/A	
Value-Based Purchasing(VBP)	No		Hospitals
Meaningful Use	No		All care providers

2.4.1 Cost–based Payment Systems

First, with cost–based payment systems, the TPP reimburses healthcare providers based on a provider's incurred costs at the time of service delivery to a patient. As this payment system only considers provider costs, the payment process involves direct flows between provider and TPP. For example, TPPs typically reimburse pharmaceutical items as a cost–based payment (Abbey, 2009). However, many TPPs no longer adopt cost–based payment systems, with the exception of some Medicare programs (e.g., Critical Access Hospitals, Rural Health Clinics (RHC), and Federally Qualified Health Centers). Instead, many TPPs require cost reports from care providers, as part of other complex payment systems such as PPSs.

2.4.2 Charge–based Payment Systems

The charge–based payment system requires TPPs to pay care providers a contractually established portion of the charged service (e.g., 80 percent of a care provider's charges). Because the TPP reimburses a percentage of whatever a care provider charges based on incurred costs, the charge–based payment system is one step removed from the cost–based payment system (Abbey, 2013). Charge–based payment systems are more likely to be associated with the use of new equipment, drugs, or procedures, driving service price increases.

2.4.3 Fee Schedule Payment Systems

Fee schedule payment systems use suitable codes to describe specific care services provided to patients. Reimbursement between a TPP and a care provider derives from a detailed service classification system of codes for physician services and items consumed (Abbey, 2011). A TPP reimburses providers for the lesser of the charge–based payment or the fee schedule amount. Current Procedural Terminology (CPT) is one classification system that includes precisely delineated physician services, such as surgical procedures, physical medicine, evaluation and management, and radiology (AMA, 2013). Based on the CPT classification, TPPs construct fee schedule payment systems. For example, the Resource–Based Relative Value Scale (RBRVS), developed by Medicare, is the basis for the Medicare Physician Fee Schedule (MPFS). Using two common classification–coding systems—CPT and Healthcare Common Procedure Coding System (HCPCS)—RBRVS applies to a wide range of services (Abbey, 2011).

2.4.4 Prospective Payment Systems

Prospective payment systems (PPS) were developed by Medicare in the 1980s (Roth and Van Dierdonck, 1995). Payments for each service are specified in advance using several unique classification systems (i.e., code sets), each of which contains specific categories (i.e., codes). A care provider charges TPPs based on categories in the applicable classification system (Abbey, 2012). PPSs cover many healthcare service types including inpatient (DRG), outpatient (APC), skilled nursing (SNF), home health (HHA), long–term care (LTCH), and rehabilitation (Inpatient Rehabilitation Facilities) services (Abbey, 2012).

The four payment systems above—cost–based, charge–based, fee schedule, and prospective payment systems—are based on the fee–for–service approach. As covered in the following subsections, other payment systems exist, such as capitated, hybrid, managed care, and operating incentive systems.

2.4.5 Capitated Payment Systems

In a capitated payment system, the payment is fixed by period (e.g., monthly), and care providers deliver any needed services without claiming additional payments within the period. This system is analogous to a theme park one–month pass, in which pass holders can ride as many rides as they want during a month. For example, some primary care clinics now offer a membership program (e.g., $75 per month) covering unlimited primary care services such as basic lab tests and flu shots (Tessman, 2014; Von Drehle, 2014). In such systems, a care provider receives the greatest financial benefit by receiving payment but minimizing service to patients during the capitation period. This potential lack of use represents a concern for TPPs and patients, as treatment underuse within a capitated approach is an expensive waste.

2.4.6 Bundled Payment Systems

In a bundled payment system, a patient encounter or series of encounters related to care provision may require bundled codes for an overall claim. Bundled payment systems offer the potential to streamline the reimbursement and payment process through predefined groupings (i.e., bundles) of services and related codes that a care provider or group of care providers offer. As bundled payments are still evolving, many issues exist regarding how to structure the bundles, design appropriate contracts, distribute payments both within and across care providers, and coordinate the bundled payment between the care provider and TPP (Hussey et al., 2009).

2.4.7 Hybrid Payment Systems

Some payment systems include more than one type of payment system. Such systems are known as hybrid payment systems. For example, for Ambulatory Surgical Centers (ASCs), Medicare developed a hybrid payment system using features of both the outpatient prospective payment system (i.e., ambulatory payment classifications) and the MPFS (Abbey, 2009).

2.4.8 Managed Care Payment System

Managed care payment systems have TPPs directly take part in the management of healthcare service. For example, the Medicare Advantage program (i.e., Medicare Part C) is an example of a managed care payment system. In Medicare Advantage, an insurance company intervenes between Medicare and Medicare beneficiaries (i.e., patients, care providers). Thus, instead of Medicare, contracted insurance companies serve as intermediaries in the reimbursement process (Abbey, 2012).

2.4.9 Operating Incentive Payment Programs

In combination with the above, TPPs such as CMS may offer healthcare providers various operating incentives to transform the quality of medical care by realigning healthcare provider financial incentives (CMS, 2014b). One example is the Value–Based Purchasing (VBP) program begun in 2012. CMS released the VBP program to connect Medicare's payment system to quality metrics (Rau, 2012). As part of the VBP program, Medicare can withhold a certain amount of reimbursements from hospitals (1% in 2012, and 2% in 2013) that do not perform well along a specified list of healthcare quality outcome metrics (CMS, 2014b). Another example is the Meaningful Use program, an Electronic Health Records (EHR) financial incentive program that provides a subsidy to care providers. With Meaningful Use, care providers adopt certified information technologies for recording patient medical information and for sharing these records (CMS, 2014a). In return for demonstrating compliance or noncompliance with Meaningful Use stages, CMS provides incentive payments or penalties (CMS, 2014a).

Admittedly, the above list of payment systems is not exhaustive, as there are other variations and hybridizations. However, the list highlights the significant variety of major healthcare reimbursement systems, which can ultimately lead to tremendous complexity for decision makers (Abbey, 2012). Yet, the number of reimbursement systems is only one antecedent of complexity. With a simple example, we next demonstrate the interaction of a basic patient presentation—a Medicare patient's laceration with reimbursement system structures, generating surprising complexity.

2.5 Illustrative Example of Reimbursement System Structure: A Laceration Presentation

At some point in life, most individuals will experience a laceration—a wound from splitting or tearing of skin. The laceration treatment process exemplifies the complexity of the interaction between a basic healthcare delivery process and complex reimbursement process structures. Figure 2.3 diagrams the laceration treatment process, including care provision (upper row, left box) and reimbursement (all other cells) involved in the patient encounter. A care provider first evaluates the severity of the patient's medical issue—in this case a laceration—and provides care. The physician then creates documentation regarding the nature of care, which flows to the billing department. The billing department evaluates and translates the physician's notes into a billing claim (see Figure 2.2).

For this example, suppose David, a 66–year–old professor, has several lacerations on his fingers and hand from working on his lawn mower. Because David has access to both Medicare (one TPP) and a university healthcare insurance plan (another TPP), he could visit a preferred provider hospital to treat his laceration (Phase 1). As the injuries are somewhat significant, David also has to meet a specialist physician several times to resolve the problem (Phase 2), followed by home health care visits (Phase 3). As Figure 2.3 shows, even though laceration treatment represents relatively simple medical care, the treatment process can generate extraordinary reimbursement complexity.

In Figure 2.3, the reimbursement process flow is represented via seven major steps. The hospital may need to use several billing systems with multiple coding layers for the laceration. While David receives various treatments for his wounded fingers and hand (Phases 1 – 3), the care providers billing system proceeds from the

Figure 2.3: Case example of Medicare patient: Life cycle of a laceration case

Care Provision: Patient-Provider

	Phase 1: Initial Presentation	Phase 2: Care Provision	Phase 3: Care Provision Post Discharge
Care Provision	**Care Provider:** General Practitioner **Action:** Escalation of Care Intensity (Referral to Specialist) **Follow-up:** Minimal Intensity After Referral	**Care Provider:** Specialist **First Action:** Decision for Inpatient or Outpatient Care. **Inpatient Action:** Hospital based care provision **Outpatient Action:** Home health plan of care through Home Health Agency (HHA)	**Care Provider:** Home health visiting nurse **Action:** Scheduled visits until termination of care

Payment Provision: Provider-Payer

	Step 4: Standard Billing Procedure from Care Provider	Step 7: Potential Audit Procedure from Care Provider
Provider Billing Systems	**Multiple Care Provider Bills:** -*General practitioner*: CPT -*Specialist*: CPT -*Inpatient Hospital*: DRG and ICD-9 diagnosis and procedure -*Outpatient Home Health Service*: HHAPPS	**Care Provider Response:** must respond to the RAC findings. Generally, the care provider must reimburse CMS. If the findings seem incorrect, the care provider must fight the charges through litigation for recovery.

Coding and Billing Provision: Provider

	Step 1: Coding Process for Initial Presentation	Step 2: Coding and Billing Processes for Care Provision	Step 3: Coding and Billing Processes for Post Discharge
Provider Coding Systems	**Billing Action:** CPT Guidelines. File 1500 claim form w/ code sets from doctor's notes	**Billing Action for Inpatient:** -*Physician*: follows CPT Guidelines. File 1500 claim form w/ code sets from doctor's notes. -*Hospital*: DRG UB04; ICD-9 diagnosis and procedure **Billing Action for Outpatient:** -*Physician*: follows CPT Guidelines. File 1500 claim form w/ code sets from doctor's notes. APC under HOPPS -*Home Health Service*: HHAPPS (see next phase)	**Billing Action:** Prospective payment system (PPS) for a Home Health Agency (HHA), which becomes an HHAPPS

	Step 5: Standard Billing Procedure from TPPs	Step 6: Potential Audit Procedure from TPPs
TPPs Payment Systems	**Medicare (CMS) Payment:** receives and pays the claims as the Third-Party Payers (TPPs): -*General practitioner*: CPT -*Specialist*: CPT -*Inpatient Hospital*: DRG and ICD-9 diagnosis and procedure -*Outpatient Home Health Service*: HHAPPS	**Medicare (CMS) Audit:** has the authority to initiate a Recovery Audit Contractor to examine all charges. Various elements are examined: -*Coding accuracy* -*Choice of care provision* (inpatient or outpatient), which alters use of the DRG or alternative billing systems.

initial presentation stage (Step 1) to care provision (Step 2) to a post discharge stage (Step 3). Though the treatment is fairly simple, each treatment step requires different billing systems under different code sets. Thus, a patient presenting with a simple laceration results in significant billing process complexity based on the decisions of the physician care provider and other stakeholders involved during each phase. Each stakeholder has differing roles in the treatment and reimbursement processes, which can lead to communication issues (e.g., translation of physician notes into billing codes) among the stakeholders. Based on the decision for inpatient or outpatient care, as well as the possibility of post–discharge home health services care, different payment systems all interact in one treatment delivery and intensify the complexity of providing appropriate care.

The post–treatment reimbursement process variability (Steps 4 to 7) amplifies the complexity. Much like a bullwhip effect in inventory systems, this variability amplification ripples across all care providers and TPPs, potentially yielding significant complexity and adverse care outcomes. Care providers and TPPs need to complete the contracted provision of payment (Step 4 and Step 5). After Step 5, CMS (or another TPP) has the power to question and dispute care provision decisions, such as the decision to treat David on an inpatient or outpatient basis. To prevent fraud and abuse (Step 6), CMS contracts with Recovery Audit Contractors (RACs) to audit care providers (AAPS, 2000). One major issue RACs investigate is the choice of site for service delivery, as CMS pays more for inpatient treatment than outpatient treatment. If RACs find that, in retrospect, the care provider used inpatient claims (i.e., the DRG payment system) too often as a means to increase charges and reimbursement rates, fines, denials, and recoupments of claims will occur. According to a nationwide tracking of 2,489 hospital systems, the number of RAC audits and the dollar amounts claimed for recovery are rising nationwide (AHA, 2013). In a single fiscal quarter of 2014, over $3 billion in recoupments and claim denials by RAC audits were reported by the 2,489 systems. Finally, care providers may need to respond to the RAC findings (Step 7).

The most common form of claim reversal reported was inappropriate use of inpatient care when outpatient care could have been sufficient (AHA, 2014)—exactly the type of scenario outlined in David's laceration. In effect, RAC auditors have authority to "second guess" the physician's choice between inpatient and outpatient care. Such second–guessing serves as a means to reverse and recoup previous reimbursements from CMS, while the RAC auditing firms collect fees based on the amount of claims recovered via audits (RAC, 2014). For each case that the RAC auditor asserts to be in error, the care provider must at least reimburse the difference between outpatient and inpatient billings (restarting Step 4 in the billing/payment loop), which can be millions of dollars for a small–to–mid–sized hospital system (AHA, 2013). In the last two years, CMS developed new regulations indicating that if an inpatient admission is not appropriate, the entire payment amount is to be recouped. If the admission is within normal claims filing guidelines, the hospital can bill for limited services. This situation illustrates how care providers must operate under financial uncertainty, not knowing whether payments from years past are final or not. Researchers might study whether physicians and hospitals might react differently if they knew payments were final after a few years.

As the example in Figure 2.3 illustrates, this situation has potential to create a serious operating incentive mismatch, as patient care quality does not appear as the prime concern for RAC auditors. When CMS uses RAC auditors to fine a hospital system and reverse claims, patient care quality may decline if the hospital system responds to reduce risk rates by enforcing lower inpatient care quality. In effect, risk of audit penalties and payment denials has the potential to override the risks to patient health. Thus, RAC auditors acting on behalf of TPPs can directly and harmfully influence patient care quality.

The reimbursement process becomes even more complex if a care provider decides to challenge RAC audit findings. Unless a mandatory dispute resolution period between the RAC auditor and care provider yields an acceptable outcome for both parties, the care provider generally pays the fines immediately or faces the risk of additional interest and penalties for non–compliance. If a discussion period fails, most cases lead to litigation in Medicare Appeals Council (Medicare, 2014), where appeals for RAC audits have a tremendous backlog, with 85% of hospitals reporting delays far in excess of 120 days (AHA, 2013). Due to persistent Medicare Appeals Council capacity issues, this backlog continues to grow, meaning the reimbursement process only becomes more complex and more costly in time and resources.

As David's example shows, various care and reimbursement processes interact during and after a patient visits a care provider. Unlike in other industries that close a reimbursement process after claim processing and payment, healthcare systems have another layer of uncertainty through processes, such as RAC audits that may occur years later to adjust and reverse claims. Even after a RAC audit reverses or recoups a claim, a provider may challenge RAC findings. In such a case, the reimbursement process may take years to finalize and involve high legal costs. Even for simple medical cases, complexity and uncertainty emerge at numerous levels over extended periods.

Unfortunately, complexities are not isolated to providers and TPPs. Patients also are affected. David may receive his own supplemental bill after an extended time, once the care provider and TPP finally determine appropriate reimbursements. Such events ultimately may affect the patients emotional and physical constitution. In effect, the complexities impact not only the care provider–to–TPP dyad, but also the entire triadic relationship.

2.6 Sources of Complexities and Uncertainties for Stakeholders

As shown, healthcare reimbursement processes drive significant operational complexity. Thus, this section delineates and describes various sources of operational complexity and uncertainty faced by stakeholders in the healthcare reimbursement processes. As in many other OM contexts affected by complexity and uncertainty (Bozarth et al., 2009; Fisher et al., 1999; Lee, 2002), healthcare operations managers need to focus on means to improve system processes. Unfortunately, complexity and uncertainty can work against such aims. Adapting concepts from prior OM/SCM studies, we define healthcare reimbursement detail complexity (we refer to this as complexity) in terms of the quantity of inputs within a healthcare system (Bozarth et al., 2009). This complexity increases as the discrete count of input quantities for a reimbursement process increases. Similarly, we define healthcare reimbursement dynamic complexity (we refer to this as uncertainty) as a reimbursement systems unpredictability of outcomes (Bozarth et al., 2009; Landro, 2013). To manage efficient and effective healthcare systems, policy makers must evaluate management of both constructs. Table 2.6 and Table 2.7 illustrate complexities and uncertainties that stem from reimbursement processes.

Table 2.6: Sources of complexity in healthcare reimbursement systems and their symptoms

Sources	**Category**	**Symptoms**
Coding Complexity	Complexity of code set systems	Increasing overhead cost for understanding code set systems Inaccurate compliance between physicians and coders
	Complexity of disease classification	Inaccurate compliance between care provider IT system and coding system Increasing overhead cost for adapting new code classification
	Complexity of care delivery team	Delays and inaccuracies of coding data transfer between care providers Inaccurate billing
Billing Complexity	Complexity of claims development procedures	Delays in filing claims Inefficient coding and billing feedback processes Care providers can be penalized by RACs
Payment Complexity	Authority of RAC to dispute any claim	Increasing additional tasks for care providers Care providers can be penalized by RACs Delays in finalizing reimbursement processes
	Complexity of rules/regulations regarding healthcare reimbursement	Increasing additional tasks for care providers Increasing payment uncertainties Reimbursement processes may not be timely finalized
	Complexity due to multi-level payers	Redundant payment adjudication processes Reimbursement processes may not be timely finalized
	Complexity of reporting systems	Delay in claims adjudication Risk of fraudulent transaction

Table 2.7: Sources of uncertainties in healthcare reimbursement systems and their symptoms

Sources	**Category**	**Symptoms**
Coding Uncertainty	Missing care information for the medical records	Delays and inaccuracies of coding data transfer between stakeholders Care providers can be penalized by RACs
	Ambiguity in the code set definitions	Inaccurate claims Inefficient coding and billing feedback processes
	Geographical access to care	Coding errors due to the different payment system between rural providers and non-rural providers Redundant reimbursement processes at different sites
Billing Uncertainty	Uncertainty of formatting of bills	Inadequate hospital claim Inefficient coding and billing feedback processes Care providers can be penalized by RACs
	Specific Requirements for various TPPs	Inaccurate bills and claims Inefficient coding and billing feedback processes
Payment Uncertainty	Timing of payment	Risk of fraudulent transaction Reimbursement processes may not be timely finalized
	Subrogation among payers	Redundant payment adjudication processes Inaccurate claim

2.6.1 Sources of Healthcare Reimbursement Complexity

To make analysis of the complexities manageable, we identify three sources of healthcare reimbursement complexities: coding complexity, billing complexity, and payment complexity.

2.6.1.1 Coding Complexity

Coding complexity pertains to coding process related input quantities. We subdivide coding complexity into complexity of code set systems for classifying procedures, complexity of disease classifications, and complexity of care delivery teams.

Complexity of Code Set System – Procedure Classification Reimbursement code sets include Current Procedure Terminology (CPT), Healthcare Common Procedure Coding System (HCPCS), and International Classification of Diseases (ICD) codes. Code sets exhibit increasing specificity as medical technology and procedures evolve. Increasing specificity forces providers to address larger numbers of codes and associated overhead from employing and educating professional coding staff. Providers must adapt to rapidly expanding coding systems. For example, CPT 2009 included 293 new codes, 133 revisions, and 92 deletions (Majerowicz, 2009), netting 201 new codes. CPT 2012 updates have 278 new codes, 138 further revisions, and 98 deletions (O'Hara, 2012), netting 180 new codes. The ever increasing code set detail complexity often drives decreasing provider compliance, resulting in incorrect payments to providers. Overpayments can drive RAC audits and other penalties against providers, while underpayments decrease profitability.

Complexity of Code Set System – Disease Classification Healthcare coding systems also increase detail complexity by including new disease classifications. For example, healthcare is moving from disease classifications based on ICD–9 to ICD–10 (Quan et al., 2008). ICD–9 and ICD–10 have completely different disease classification formats. Such classifier incompatibilities dramatically increase the number of codes in a billing package and the risk of non–compliance. Thus, expanding disease classification is one source of growing coding system complexity.

Complexity of Care Delivery Team Certain payment systems require only one payment for all services involved in a medical procedure. Multiple care providers can take part in care provision, but billing takes place as a single, unified billing package. This package tends to require numerous codes, increasing likelihood of errors or omissions. For instance, a global surgical package (GSP) contains multiple providers, including a surgeon, anesthesiologist, and an operating room facility. Surgeries take place as inpatient or outpatient, and usually entail pre–operative and post–operative protocols, which may be provided by other physicians or other qualified healthcare personnel. Although the unified billing package should streamline the process, the large number of providers increases coding detail complexity, which complicates the billing process.

2.6.1.2 Billing Complexity

Billing complexity pertains to billing process related input quantities. Specifically, billing complexity deals with complexity of claims development procedures.

Complexity of Claims Development Procedures Given payment code set complexities, care providers confront difficulties when preparing accurate and timely billing claims. Care providers generally must claim and file reimbursement requests within a year of healthcare service provision (Tricare, 2013). The federal Health Insurance Portability and Accountability Act of 1996 Transaction and Code Standard (HIPAA TCS) provides billing process standards for care providers (CMS, 2013). Providers also must meet other requirements from TPPs (i.e., CMS or insurance companies) to file claims. In some cases, providers must file claims with TPPs about which they have no knowledge or guidance. The expanding nature of HIPAA and TPP adjudication standards results in detail complexity, delaying claims filing and increasing inaccuracies.

2.6.1.3 Payment Complexity

Payment complexity pertains to payment process related input quantities. Payment complexity is driven by processes, such as the RAC authority to dispute reimbursement claims, statutory payment system regulations, multi–level payers, and reporting systems.

Complexity Due to RAC Authority to Dispute Claims The role of RACs, which audit care providers to identify and correct improper payment processes, increases payment detail complexity. A RAC has authority to reverse claims and question or second–guess previous care provision or service encounter choices. Second–guessing increases payment complexity because RAC reexamination may occur long after the completion of care provision. With more ways to dispute claims, professional coding staff must prepare for alternative outcomes (e.g., denials, disputes, or partial payments) that increase payment complexity. In effect, the RAC can create short–term and long–term complexities for the reimbursement process, because of the increasing number of tasks required to improve the likelihood of receiving and retaining a payment.

Complexity of Regulations in Statutory Payment Systems To keep up with expanding code set systems and billing procedures, statutory payment rules and regulations also continue to increase (Field, 2008). For instance, a care provider must follow a specific payment protocol to meet Medicare rules. As this payment protocol expands, care providers face added bureaucratic demands and contracts with private insurers that further increase payment complexity.

Complexity Due to Multi–Level Payers Patients and providers may contract with multiple TPPs. The use of multiple TPPs increases payment system complexity. HIPAA, which mandates that all claims for a given service must be standardized, was designed to enhance the processing of claims through multiple TPPs. In reality, however, providers still experience significant payment complexity due to wide disparity of payment systems among primary, secondary, and tertiary TPPs. For instance, the Medicare Secondary Payer (MSP) program requires a full standalone CMS manual for clarification of its programs. As a result, the obvious holds—as payment systems include more secondary and tertiary payers, payment complexity will rise.

Complexity of Reporting Systems In keeping with secondary payment processes, Medicare developed a complex reporting process for primary payers to clarify liabilities. Intricate reporting systems, such as MMSEA Section 111 (Edelson, 2013), were developed to address the role of primary payers in multi–payer payment systems. Additional layers of reporting add more tasks to the overall payment system, increasing payment detail complexity.

2.6.2 Sources of Healthcare Reimbursement Uncertainties

Healthcare reimbursement systems also embed many operational uncertainties, exposing stakeholders to coding, billing, and payment uncertainties.

2.6.2.1 Coding Uncertainty

Coding uncertainty relates to coding procedure unpredictability. Coding uncertainty arises from missing care information, ambiguity in code set definitions, issues with translation of physician notes into codes, and geographical care provider heterogeneity.

Uncertainty from Missing Care Information in Medical Records In general, physicians follow a care providers protocol for documentation. Though a physician might have some idea of the codes used within a billing system, most physicians do not have comprehensive knowledge of the coding systems. A lack of knowledge generates variation in the coding process (Powell et al., 2012). Ignoring coding process variation can increase reimbursement uncertainties and exacerbate audit risk. Coding process variation starts from characteristics of care procedure notes that can be interpreted to be different types of services (e.g., emergency admission with trauma department vs. without trauma department). In the end, frontline employee (i.e., physician) uncertainty leads to coding process variation, causing conflict among physicians, professional coding staff, and TPPs, as each holds different incentives for reimbursement outcomes.

Uncertainty from Ambiguity in Code Set Definitions and Use Under HIPPA, healthcare providers and TPPs must use standard code sets in reimbursement processes. However, dozens of different standard code sets exist, along with multiple organizations that maintain the code sets. For instance, a care provider should use at least two procedure codes for an ER laceration treatment, one for evaluation and one for the procedure. However, this care treatment claim might be delivered with only one bundled code set, leading to non–compliance. Though not intentional, provider organizations may not adhere to proper use of standard code sets and formats. Also, some TPPs prefer to use only subsets of available codes—another non–compliance. Thus, variable application of code set requirements, due to existence of many code set standards, increases coding uncertainty.

Uncertainty from Geographical Access to Care Geographical dispersion of care providers delimits access to care and requires application of different code sets. To address geographical dispersion, TPPs adopt different payment systems to promote efficient care delivery in some regions, with a lesser focus on effective care delivery. Thus, rural healthcare providers, such as Critical Access Hospitals, RHCs, and Federally Qualified Health Centers, use delimited code sets under the Medicare program to support efficiency and viability of the organizations through narrower ranges of covered procedures (CMS, 2015b). Overall, geographical differences in access to care can increase uncertainty in the appropriate coding systems.

2.6.2.2 Billing Uncertainty

Billing uncertainty pertains to dynamic complexity related to the billing process. Billing uncertainties can be caused by bill formatting and specific TPP requirements.

Uncertainty of Formatting of Bills Based on physician documentation, professional coding staff creates bills for reimbursement (see Figure 2.2). The significant role of professional coding staff of a care provider is appropriate bill creation for submission to external TPPs for payment. The coding staff relies on the already variable documentation of physicians. Coding staff often faces ambiguous codes and difficulty translating physician notes and medical records into specific codes (Powell et al., 2012). Due to the coding process variation, coding staff has various options when creating the billing claim for TPPs. Variable bill formatting represents a source of billing uncertainty.

Uncertainty from Specific Requirements of Various TPPs Billing uncertainty also arises from specific requirements from various TPPs. For example, Medicaid and Medicare have different demographic criteria for participants. Medicaid program requirements assist people with low income, while Medicare requirements assist people 65 years of age or older (CMS, 2015b). In addition, private insurers contractually limit the range of covered services. This limited range of covered services is not always clear to patients or care providers until a claim has been processed. Additionally, coverage may be delimited by assertions that services are not medically necessary. Variability in specific requirements within and across TPPs for a patient can amplify billing uncertainty.

2.6.2.3 Payment Uncertainty

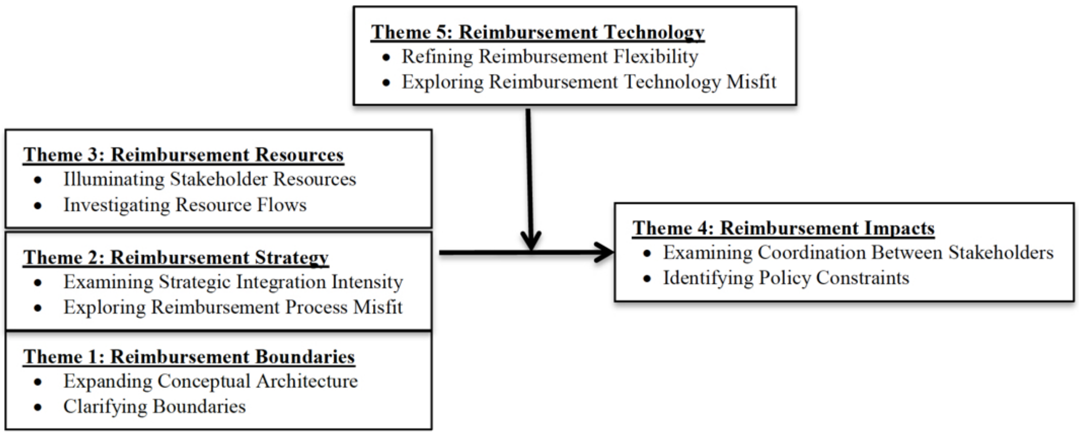
Payment uncertainty pertains to payment process related unpredictability. Payment uncertainty results from timing of payment and the potential for subrogation among payers.

Uncertainty in Timing of Payment One source of payment uncertainty concerns the payment finality problem. Some states have requirements that all transactions be closed within a year, or some other specified period, after the first claim. However, a TPP may reverse and recoup payments on claims after many years if the TPP finds evidence of fraudulent or faulty claims during a later audit. Such recoupments are particularly common for CMS, which uses RAC audits. This issue relates to both coordination among parties and the time before finality of payment—both issues that potentially impact provision of healthcare and reimbursement processes. Thus, variability in payment finality increases payment uncertainty.

Uncertainty Due to Subrogation Among Payers TPP subrogation raises uncertainty in reimbursement processes. Subrogation is when one TPP takes over payment obligations from another TPP. Although HIPAA exists to standardize all reimbursement practices, subrogation among payers can lead to increasingly uncertain payments due to variations in claim reimbursement and adjudication as billing claims traverse a chain of secondary and tertiary TPPs.

In summary, this section outlines several sources of complexity and uncertainty in healthcare reimbursement. The list suggests the potential for impactful OM/SCM research to better understand and improve healthcare operations. Based on the preceding conceptual frameworks, the next section details research opportunities pertaining to healthcare reimbursement processes.

Figure 2.4: Inter-relationship between research themes



2.7 Implications for Research Opportunities

Having established conceptual foundations for reimbursement processes, we turn toward the principal focus of this essay: unexplored reimbursement research opportunities in healthcare OM/SCM. This section endeavors to reduce the seeming chaos of reimbursement complexity and uncertainty into five research themes. Figure 2.4 shows inter–relationships among the research themes.

2.7.1 Expanding Healthcare Operations Boundaries to Include Reimbursement Processes

To better address healthcare management issues, scholars need to expand research boundaries to include operations challenges generated by reimbursement processes. Studying service operations requires the integration of multiple disciplines including operations, human resources, marketing, and finance (Boudreau et al., 2003; Roth and Menor, 2003). Studying contemporary healthcare will also demand integrative multi–disciplinary, multi–methodology work. Expansion of research boundaries to include reimbursement processes will be of particular benefit to service OM literature, which rarely considers reimbursement processes or broader financial operations. Examining enlarged healthcare functional boundaries, by including a TPP, could shed light on the ambiguity of current healthcare reimbursement processes and unexpected consequences for service delivery. In summary, expanding functional research boundaries to include TPPs is important for considering the impacts of financial transactions upon healthcare services.

2.7.1.1 Expanding Conceptual Healthcare Architecture Boundaries

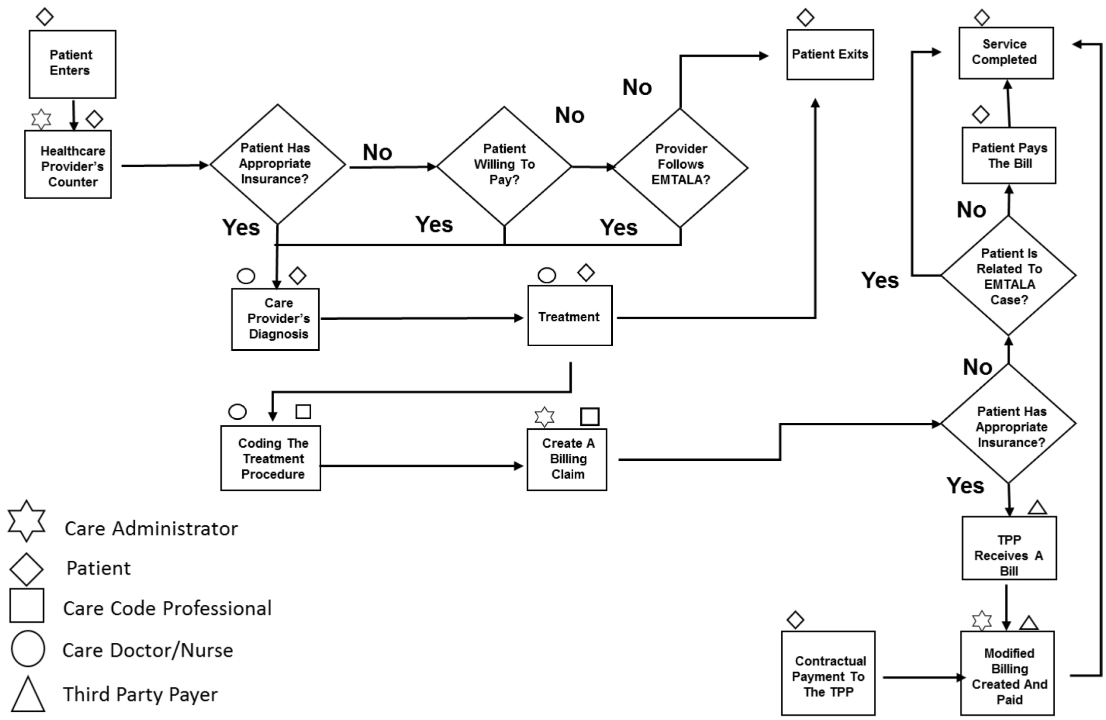
As with the boundary expansion of service concepts into traditional OM (Heskett, 1986; Roth and Menor, 2003; Voss, 1992), conceptualizing the full architecture of stakeholders in healthcare delivery will expand healthcare OM boundaries. This essay provides a critical first step of detailing conceptual foundations. Future research should expand healthcare architecture boundaries to consider how reimbursement processes serve as a major driver of the overall healthcare process. Thus, we posit:

RO 1: Conceptualizing the full healthcare delivery architecture with expanded boundaries including reimbursement processes.

2.7.1.2 Clarifying Boundaries of Reimbursement Systems

Healthcare OM researchers often make simplifying assumptions about reimbursement systems applicable to a study (i.e., DRG is assumed). Researchers should clarify which set of reimbursement systems apply to a research study and how the systems mediate or moderate quality of care outcomes. For example, Figure 2.5 shows healthcare reimbursement resource flows processed by five stakeholders, such as a care administrator, patient, code professional, doctor, and third party payer. Doing so may force researchers to explicitly diagram care and reimbursement resource flows, as illustrated in Figure 2.5, to identify relevant system boundaries. Researchers in healthcare OM/SCM have not yet recognized that exploring the differential im-

Figure 2.5: Healthcare reimbursement resource flow



pacts of diverse reimbursement processes has an equally powerful influence to those of the often–studied clinical treatment processes. The healthcare system under study, which should include reimbursement processes, represents a further area of study on how to develop effective and efficient infrastructure. Thus, we posit:

RO 2: Clarifying the overall scope of healthcare reimbursement systems to uncover how each reimbursement system influences care provision.

2.7.2 Refining Operations Strategy for Healthcare Reimbursement

Researchers must enhance OM/SCM theory regarding reimbursement processes. Through theory, a statement of the nature of relationships among constructs (Amundson, 1998; Handfield and Melnyk, 1998), healthcare scholars can deepen understanding of the nature of healthcare reimbursement processes. When boundaries of an

OM research area grow, the need for strategy research also grows. Thus, our second theme concerns refining the field's understanding of healthcare operations strategies. As with prior emerging research areas in service OM that focused on new topics such as e–retailing (Heim and Sinha, 2002) and new service development (Hill et al., 2003), healthcare scholars now need to broaden and refine operations strategy theories to facilitate examination of new healthcare factors that are yet to be studied. This section presents several research opportunities pertaining to healthcare operations strategy.

2.7.2.1 Strategic Integration Intensity of Healthcare Service Delivery

In manufacturing OM/SCM literature, the degree of integration intensity is positively associated with operational capabilities and knowledge assets (Rosenzweig et al., 2003; Swink et al., 2007). Yet, integration intensity may not provide expected outcomes if an organization does not have needed capabilities (Devaraj et al., 2007). Similarly, healthcare OM/SCM literature needs to consider concepts of strategic intensity. Healthcare integration resources need to support both patient care delivery encounters and reimbursement processes. To understand healthcare integration intensity, researchers should understand patient encounter intensity and care provider reimbursement intensity. Studying various strategic dimensions of healthcare system intensity will help researchers and practitioners better understand the full scope of healthcare strategic management. Researchers will need to define and examine systems integration intensity, patient encounter intensity, and reimbursement intensity and its sub–constructs such as coding, billing, and payment intensity. Table 2.8 illustrates the scope of patient encounter intensity and care provider reimbursement intensity. Appendix B provides detailed descriptions of stages within each intensity dimension, providing initial work on healthcare system intensity. More rigorous

Table 2.8: Healthcare system intensity dimensions

Diagnosis	**Patient Encounter Intensity**	**Definition:** The severity level of a patient's risk of complications, morbidity, or mortality.	
	Levels	Examples	Complexity of Care Provider Care Decision
	Minimal intensity	Self-care patients, preventive care	Straightforward
	Low intensity	Short encounter duration, outpatient, home care	Low Complexity
	Moderate intensity	Middle encounter duration, outpatient	Moderate Complexity
	High intensity	Emergency patients, OR care	High Complexity
Recording Documenting	**Care Provider Reimbursement Intensity**	**Definition:** The amount of data and complexity of medical records to be reviewed, coded, billed, processed for payment, and audited.	
	Levels	Examples	Complexity of Care Provider Care Decision
	Minimal intensity	Self-care patients, preventive care	Straightforward
	Low intensity	Short encounter duration, outpatient, home care	Low Complexity
	Moderate intensity	Middle encounter duration, outpatient	Moderate Complexity
	High intensity	Emergency patients, OR care	High Complexity

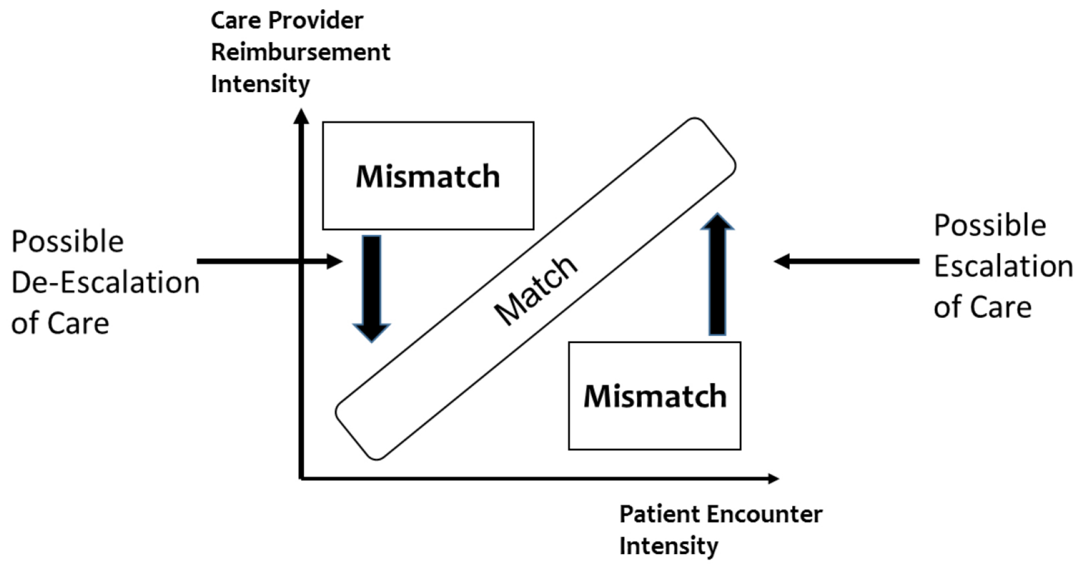
studies, such as analytical models of the value of care provider integration intensity, should clarify specific reimbursement process issues. Therefore, we state:

RO3: Understanding and investigating healthcare strategic intensity impacts on care delivery and financial outcomes

2.7.2.2 Strategic Reimbursement Misfit

Strategic misfits occur between patient needs and care provider capabilities. Figure 2.6 diagrams the relationship between patient encounter intensity and care provider reimbursement intensity. When an individual presents as a patient, the patient may independently choose a care provider based on personal utility (e.g.,

Figure 2.6: Misfit between patient intensity and care provider intensity



wealth, prior experience, preferred provider status, or proximity). A patient choice may exhibit high or low fit with provider care and reimbursement capabilities. The patient vs. care provider structure in Figure 2.6 is similar to extant service operations frameworks. Just as intensive technology enables expert service processes (Kellogg and Nie, 1995), high intensity providers can deliver more appropriate services for a high intensity patient requiring greater healthcare resource levels (Jencks and Dobson, 1987).

Two types of mismatch can occur in matching patient encounter intensity to care provider reimbursement intensity. The first type occurs when patient encounter intensity is high, but care provider intensity is low. Patients may consider their symptoms as fitting the low intensity level, so they present to a low intensity care provider. However, if the patient intensity level is actually higher than assumed, a mismatch occurs. This mismatch is typically resolved through escalation, wherein a care provider who encounters a mismatched case escalates the patient to a higher intensity care provider (e.g., from clinic to hospital). Thus, a high patient intensity/low care provider intensity mismatch can self–correct through escalation toward the matched area. The second mismatch type occurs when a low intensity patient presents to a high intensity care provider. For example, a patient with a Grade–1 finger sprain (low patient encounter intensity) may present to a hospital emergency room (high provider intensity). This mismatch might be resolved by de–escalating care (e.g., from hospital to clinic). However, as the high intensity care provider has the capability to treat this patient, the mismatch generally will not be resolved, and the patient often remains at the high intensity provider. A *Wall Street Journal* article (David, 2014) highlights this mismatch type with an example of a patient with a minor bruise (i.e., low patient encounter intensity) who chose to use a trauma center (i.e., high care provider intensity), leading to an excessively high $20,000 charge. Less severe mismatches are common in hospital ERs in which non–emergent care is provided. Thus, we posit:

RO 4: Exploring the impact of reimbursement intensity levels on care outcomes and financial flows using misfit, alignment, and complementarity theories.

2.7.3 Illuminating Healthcare Reimbursement Resources

Researchers should investigate healthcare reimbursement resources that have not been considered. Healthcare providers and TPPs have been implementing many new resources, such as Electronic Medical Records (EMR), Health Information Exchanges (HIE), and Personal Web Portals (PWP). Strategic management theories, such as institutional theory (Bhakoo and Choi, 2013), the relational view (Chen et al., 2013), the theory of swift and even flow (Devaraj et al., 2013), the theory of absorptive capacity (Boyer et al., 2012), and organizational learning theory (Ding, 2014), have been used to highlight key care–oriented constructs of healthcare OM/SCM contexts. Researchers should also examine reimbursement processes using the above theories, as well as theories derived from process management and management of technology perspectives.

2.7.3.1 Reimbursement Stakeholder Resources

Researchers may use seminal theories to explore phenomena motivating reimbursement processes in healthcare. Studies on resource dependence theory explore links among organizations as a set of power relations based on resource interchange (Goes and Park, 1997; Hillman and Dalziel, 2003; Pfeffer and Salancik, 2003). Healthcare organizations depend greatly on external reimbursement related resources, such as government capital inputs (i.e., Medicare or Medicaid), regulations and legislation (i.e., codes, RACs, or HIPAA), and labor inputs (i.e., doctors, nurses, administrators, or professional coders). Each organization has different social capital. Thus, an organizations provision of reimbursement resources and social capital disparity can lead to different power relationships between organizations. Much extant work in OM/SCM clarified the nature of supply chain relationships (Flynn et al., 2010; Koufteros et al., 2007), thus the healthcare domain also necessitates assessing and clarifying the nature of stakeholders in the healthcare reimbursement system. Thus, we propose:

RO 5: Clarifying the nature of stakeholder resources in the healthcare reimbursement system.

2.7.3.2 Reimbursement Resource Flows

Based on the above insights, conceptual healthcare reimbursement resource flow models need to be developed for healthcare OM/SCM research to fill in theoretical gaps between academic research and healthcare practice. Just as there exist many coordination models in SCM with different optimal policies (Cachon, 2003), healthcare reimbursement resource models should be constructed to analyze different coordination conditions for maximizing healthcare system performance. Using common agency theory (Rebitzer, 2014), we illustrate one example of reimbursement resource (i.e., incentives). When a large TPP (i.e., CMS) commits to incentive contracts with a care provider, other TPPs may also seek to contract with the care provider, particularly if the other TPPs offer similar contracts to that offered by the large TPP. Thus, we posit:

RO 6: Designing conceptual healthcare reimbursement resource flow models to enable analysis of complex real–world systems

2.7.4 Examining Reimbursement Process Impacts on Stakeholders

Researchers should study the multidimensional impacts of reimbursement in the healthcare sector upon stakeholders. Evidence suggests patients can be emotionally harmed due to billing errors, and potentially physically harmed if such errors adversely affect a patients treatment decision making (David, 2014). Reimbursement process errors can lead to higher costs to patients, lower patient credit scores if mistaken charges occur (Silver-Greenberg, 2011), and increased medical collections (Bernard, 2012). When billing errors occur, TPPs must process and reimburse more bills to the healthcare providers, and in response, TPPs often impose penalties upon the providers. Addressing error–laden activities represents excess overhead costs that are unnecessary and avoidable. To date, healthcare OM/SCM researchers appear to have perceived that exploring processes pertaining to reimbursement may be less substantial than other medical issues, such as clinical treatment errors. However, reimbursement process investments are substantial, and their consequences are not negligible. Therefore, the varied impacts of reimbursement process should be studied.

2.7.4.1 Coordination Among Healthcare Stakeholders

Each stakeholder—patient, provider, and TPP—plays different roles within the healthcare reimbursement system. Without proper coordination within and among healthcare system members, the financial healthcare system will not operate effectively (Frandsen et al., 2015). Inefficient coordination can decrease stability of healthcare reimbursement. In particular, missing or ineffective information sharing among stakeholders can contribute to inefficient coordination. Brill (2015) suggests aligning stakeholder incentives by reforming healthcare payment systems, merging providers with TPPs, and imposing strong regulations. Contracts between stakeholders or incentive issues also can cause inefficient coordination and inefficient investment of healthcare system members. Common agency theory might be adopted to examine care providers incentives, expenditures, and coordination through the collective action of stakeholders contracts (Frandsen et al., 2015). To date, relationships among healthcare stakeholders in terms of reimbursement processes have not been explored. As such, we posit:

RO 7: Investigating each stakeholders behaviors and incentives in the healthcare financial system.

2.7.4.2 Identifying Policy Constraints on Stakeholders

The healthcare domain strongly depends on external constraints (Bhakoo and Choi, 2013; Chandrasekaran et al., 2012). In organizational strategy literature, institutional theorists stress that external constraints, such as government regulation, influence stakeholder's behavior (DiMaggio and Powell, 1983; Scott et al., 2000). Though some previous healthcare OM studies examine impacts of external pressures on care delivery (Bhakoo and Choi, 2013; Chandrasekaran et al., 2012), few studies examine external constraints forced upon healthcare reimbursement systems, such as Meaningful Use and Value Based Purchasing (VBP) initiatives. Therefore, we posit:

RO 8: Exploring impacts of regulatory policy, external constraints, and pressures upon healthcare reimbursements for healthcare patients, providers, and payers.

2.7.5 Understanding Impacts of Reimbursement Technology

The final theme relates to impacts of reimbursement technology. A care provider may adopt reimbursement technology in the form of various payment modules and systems, depending on provider characteristics and capabilities. These payment systems often must integrate to internal and external healthcare delivery technology such as EHR, HIE, PWP, and mobile devices (HealthIT, 2014). We define reimbursement flexibility as the degree of reimbursement options a stakeholder can use for reimbursement processes. Examining reimbursement flexibility should provide insights for OM researchers to better understand healthcare financial systems. Figure 2.7 expands off of Figure 2.1 to build a more comprehensive model of IT enabled flows.

2.7.5.1 Technology Reimbursement System Flexibility

Recent evidence suggests the development of innovative technology for payment systems has grown in the healthcare industry to enable better healthcare system operations and improved care quality (Abbey, 2009; Rosenthal, 2008). Yet, such claims are open to examination. For example, Medicare annually updates to improve payment processes for patients, care providers, and TPPs (OIG, 2014). With the expansion of different types of payment systems, a care provider has an expanding list of different types of payment challenges depending on its health care service types. If a care provider offers several types of services, such as inpatient, outpatient,

Figure 2.7: IT enabled healthcare service delivery system flows

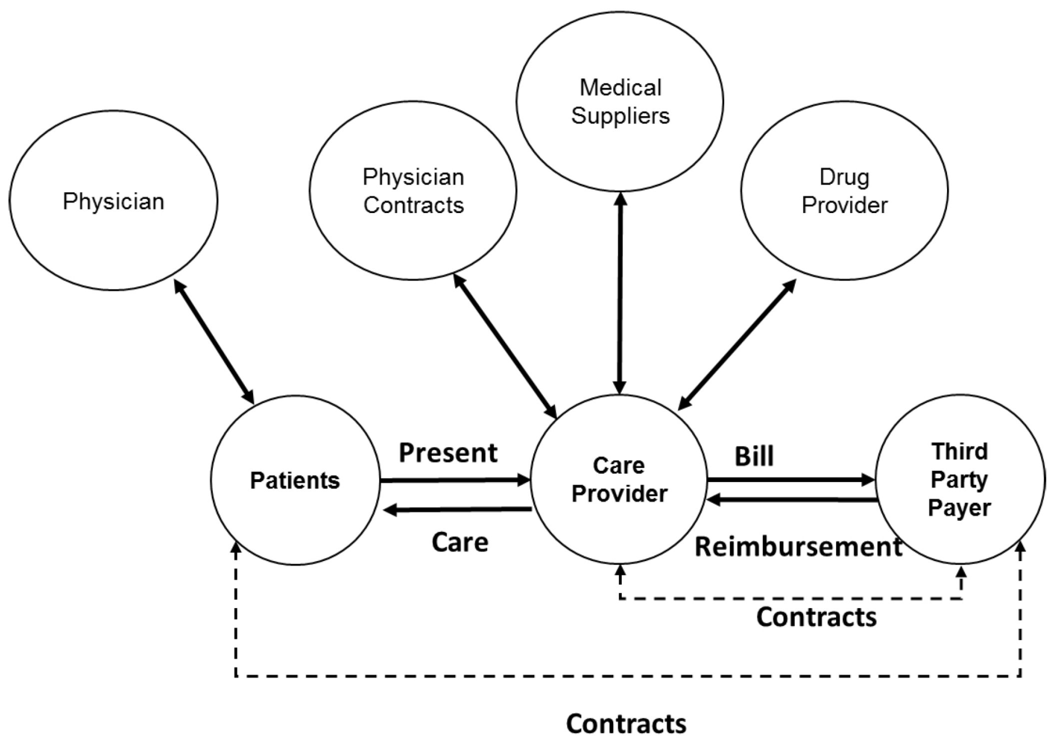
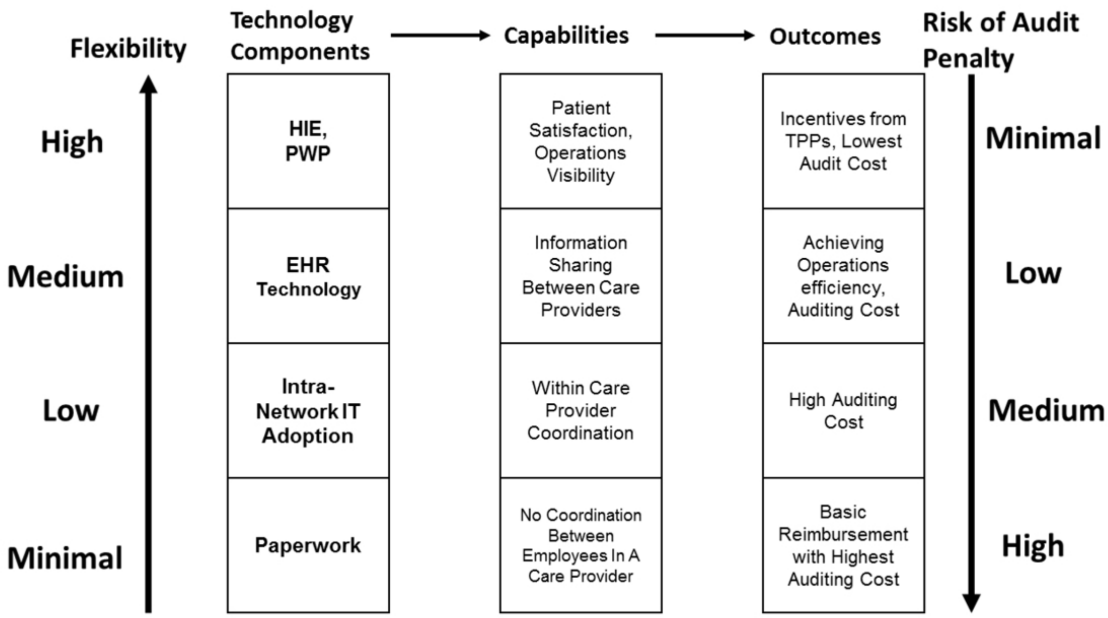


Figure 2.8: Impacts of reimbursement flexibility



and emergency care services, it may have to accommodate fee schedule payment systems, prospective payment systems, capitation payment systems, and contractual payment systems. Thus, the care provider technology must have a high degree of reimbursement flexibility, as the care provider must accommodate different payment methodologies from multiple TPPs. Figure 2.8 shows a diagram of reimbursement flexibility, which we suggest will have related technology components, care provider capabilities, and corresponding outcomes. Research should examine the role of technology relative to reimbursement flexibility:

RO 9: Investigating the role of technology in supporting reimbursement flexibility.

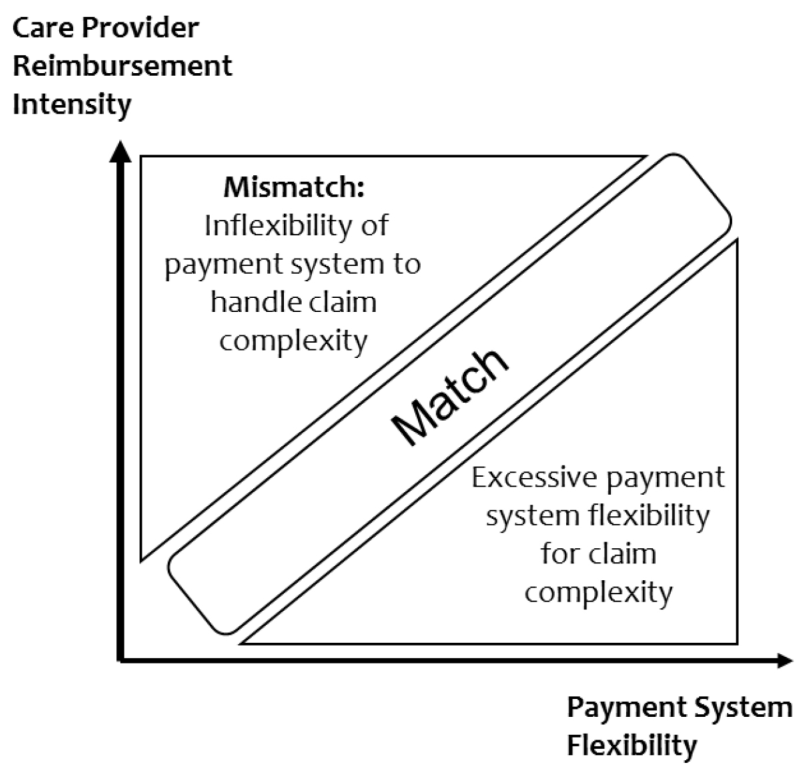
When a patient is discharged from a care provider, the provider contacts the corresponding TPP responsible for payment. As patient encounter intensity increases, the necessary care provider intensity increases (Jencks and Dobson, 1987), because more treatment procedures will be required. As the increasing number of treatment procedures becomes complicated, the way in which providers are reimbursed becomes more varied and complex. Thus, the payment systems in a high intensity care provider must have flexibility to be paid based on an individual patient's varying payment requirements. When a mismatch exists in the care delivery phase, a mismatch often follows in the reimbursement processing phase (Figure 2.9). If the professional coding expert responsible for a claim faces an unexpected unsuitable match between patient and care provider under ambiguous coding conditions, the coding is likely to be inaccurate, which can result in medical billing errors. In addition, when the mismatch happens frequently, coding experts are more likely to use self–preferred codes, which may not be appropriately matched with the patient's actual procedure. In some cases, unlisted general codes can be used, which then affect reimbursement. Thus, we propose:

RO 10: Exploring the impact of care delivery misfit upon care outcomes and financial flows via reimbursement technology misfit

2.8 Conclusion

This essay highlights healthcare operations processes that have yet to be examined by healthcare OM/SCM scholars. We first demonstrate the roles and complexities of healthcare reimbursement processes. Subsequently, we identify dimensions of complexity and uncertainty experienced by healthcare stakeholders. We finally suggest unexplored research opportunities. Given the complexity of healthcare reimbursement problems, the lack of research appears to be at least partially driven by the lack of an appropriate research agenda. Based on our agenda of research opportunities, researchers may make substantial contributions by extending their managerial interest in OM/SCM topics to healthcare reimbursement processes. The

Figure 2.9: Billing system: Care provider match vs. mismatch

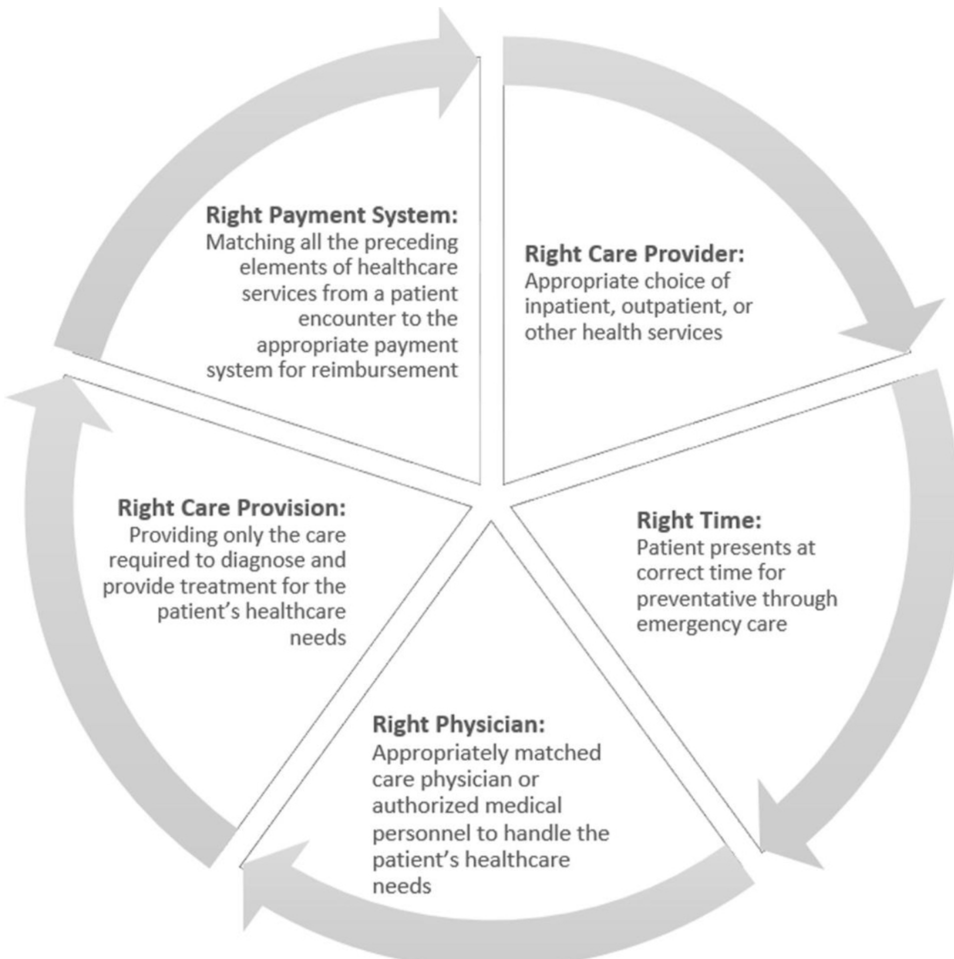


frameworks and research opportunities outlined throughout this essay are a major first step for the field of OM/SCM to address the Five Rights (5Rs) of a healthcare supply chain. Specifically, as shown in Figure 2.10, the 5Rs of a healthcare supply chain are: right care provider, right time, right physician, right care provision, and right payment system. These 5Rs play a crucial role in managing a patient's medical needs by matching all the various elements of the healthcare supply chain from the choice of the care provider through the eventual reimbursement via a payment system. If research continues to fail to account for the significant roles of reimbursement processes, scholars will fail to account for all factors that enable high quality healthcare service—a case of omission bias.

Due to the broad set of characteristics in modern healthcare, our suggested research opportunities provide a first step and are admittedly not exhaustive. Yet, based on our review of related literature, reimbursement research opportunities should gain increased attention and become increasingly important problem domains for several reasons, including the sheer scope of healthcare reimbursement as a driver of GDP. While some argue new healthcare laws have improved care access and reimbursement (HHS, 2015), contemporary healthcare reimbursement trends suggest reimbursement managers will need to consider rapidly evolving technology choices, consolidation for operational and reimbursement economies of scale, even more reimbursement models, and how insurance exchanges will affect stakeholders (Weldon, 2014). Healthcare service researchers hold a rare opportunity to be among the first to examine healthcare reimbursement processes and related operational issues. To this end, the essay provides initial directions for future studies into healthcare operations and outlines the need to account for the 5Rs of a healthcare supply chain.

Relatedly, this study also offers high–level guidance for practitioners. Our conceptual frameworks and constructs may help healthcare managers identify operational

Figure 2.10: The Five Rights of a healthcare supply chain



antecedents of administrative issues, which may enable reduction of overhead and administrative error costs. If academic researchers can provide constructive research questions and corresponding prescriptive advice, healthcare will be further improved in terms of efficient and effective operations management. We believe our frameworks related to healthcare reimbursement processes and corresponding research opportunities will enable appropriate questions regarding unexplored reimbursement issues for the OM/SCM field.

3. OUTPATIENT APPOINTMENT SCHEDULING UNDER PATIENT NO-SHOWS AND PATIENT HETEROGENEITY

3.1 Introduction

Healthcare providers face many challenges in improving efficiency and effectiveness of health care systems (Cayirli et al., 2006). For outpatient clinics, employing a well-performing appointment scheduling policy is critical to balance the needs for efficient resource utilization (by mitigating physician's idle time and overtime) and effective care delivery (by minimizing patients' waiting time to enhance satisfaction). Appointment scheduling research has provided many useful scheduling policies for outpatient services (Cayirli and Veral, 2003). Nevertheless, outpatient scheduling systems need to be studied further to account for even more realistic issues (Cayirli et al., 2006; Chen and Robinson, 2014).

Recent events illustrate potential shortcomings of outpatient scheduling systems used by practitioners. For example, in response to manager performance incentives based on outpatient scheduling targets, Veterans Affairs (VA) hospitals confronted a serious scandal due to falsified outpatient scheduling data and poor care outcomes (Kensling and Nissenbaum, 2014). More broadly, outpatient physicians today feel overwhelmed by the complexity of scheduling their practices due to the extent of patient forgetfulness and diverse patient types requiring different care and service times (Majd, 2015). With recent developments in healthcare IT, many patients soon will be asked to use web-based appointment scheduling systems to self-schedule outpatient appointments (Williams, 2015). However, due to the high costs of implementing advanced IT for scheduling, many outpatient clinics can be financially constrained. Thus, there is an urgent need to develop simple but effective scheduling approaches for such practitioners. Compounding such issues, U.S. outpatient expenditures recently have grown by 10% annually — double the rate of inpatient services — due to expansion in aging populations, chronic diseases, accountable care initiatives, and pay-for-performance reimbursement pressures (HCCI, 2010). These many challenges for outpatient healthcare services motivate a need to further examine outpatient scheduling to identify realistic and flexible scheduling policies.

3.1.1 Two Challenges

In this study, we focus on two salient challenges: patient heterogeneity and patient no-shows. Patient heterogeneity and patient no-shows can lead to inefficient and ineffective scheduling outcomes (Gupta and Denton, 2008), yet little academic research considers these two factors together.

Appointment schedule performance can be sensitive to patient heterogeneity, due to patient gender, age, physiology (Cayirli et al., 2006; Majd, 2015), and new or follow-up patient types (AMA, 2013, p.11, Chapter 1). Literature suggests the main difference between two types concerns required service times (Cayirli et al., 2012). Since patient heterogeneity exists, scheduling algorithms should take this information into account when developing patient schedules (Chen and Robinson, 2014). However, no one specific appointment policy can outperform all other policies due to each clinic's different environment (Cayirli et al., 2012). Further, most outpatient scheduling rules are analytically intractable, making it difficult to demonstrate an algorithm's optimality (Gupta and Wang, 2011). Thus, instead of pursuing a universally optimal approach, we propose an adaptable sequential block scheduling policy that reflects the way many outpatient care providers prefer to incorporate patient heterogeneity, that is, via heterogeneity in service times.

Patient no-shows, where a patient does not show up for an appointment, also can harm operational efficiency, through under-utilization of outpatient service system capacity (Zacharias and Pinedo, 2014). Typical outpatient appointment scheduling approaches allow patients to make appointments weeks or months in advance, leading to patient no-shows due to forgetting about appointments or scheduling conflicts (Chen and Robinson, 2014). Previous studies examine impacts of no-shows on financial performance (Moore et al., 2001; Pesata et al., 1999) and provider-patient relationships (Pesata et al., 1999). Literature also finds patient no-show rates vary across clinics (Robinson and Chen, 2010) and patient heterogeneity and no-show rates are interrelated (Liu and Ziya, 2014).

While prior literature explores patient heterogeneity and patient no-shows separately, not much literature considers both together (Zacharias and Pinedo, 2014). The objective of this study is to provide new scheduling approaches for outpatient facility managers who face patient heterogeneity in service times and patient no-shows. In line with extant literature, we consider three concerns affecting the design of an appointment system. First, clinics must consider patients' waiting time. Second, clinics may try to avoid overtime of physicians. Third, clinics may try to minimize physician idle time. This study tries to balance the three concerns.

3.1.2 Our Approach

We develop sequential block scheduling procedures inspired by the load smoothing methods used in the Toyota production system. Our appointment scheduling approach incorporates several assumptions. First, following extant literature (Robinson and Chen, 2010), we assume that a single physician represents one care delivery system, which accounts for the one-to-one doctor-patient relationship. In reality, patients prefer to seek services from the physician of their choice. Thus, a patient scheduled to visit one particular physician cannot be transfered to other physicians who are underutilized.

Second, we assume a care provider knows each patient's information in advance, making it possible to assign a type to each patient well ahead of time. In healthcare practice and academic literature (Chen and Robinson, 2014; Gupta and Denton, 2008), the assumption that demand for service by patients is large enough to consistently fill appointment slots within each day is well-accepted. In doing so, we can assign a fixed number of patients from different types into the future time slots. In addition, one main reason for patient no-shows may be a patient's forgetfulness due to a clinic scheduling appointments well ahead of time. Thus, the second assumption motivates the need to investigate patient no-shows.

Third, we assume that the scheduled patient arrivals are fixed. The fixed time slot appointment is well accepted in modern outpatient clinics (Huang and Hanauer, 2014), and the academic literature (Cayirli and Veral, 2003; Mondschein and Weintraub, 2003; Schütz and Kolisch, 2013) also provides evidence that the fixed time slot assumption is not uncommon in outpatient appointment scheduling practice.

Based on the above assumptions, the study develops block scheduling policies for several scenarios: zero no-show, positive no-show, overbooking, and open access.

3.1.2.1 Block Scheduling Policy

Each block scheduling policy is inspired by the load smoothing methods used in the Toyota Production System (i.e., "Heijunka") (Ono, 1988), which tries to balance different demand types and production volumes. We adapt the Toyota concepts to assist outpatient scheduling managers who face patient heterogeneity and patient no-shows. In particular, our policy tries to convert different patient service times and numbers of patients for multiple patient types into an evenly balanced and predictable process to stabilize physician idle time and overtime. In essence, each patient type becomes an input (comparable to a product type in Heijunka) that a facility (i.e., physician) has to serve within a planning horizon. Our block scheduling policy requires the following steps: (1) patient types are identified, (2) information on relative demand frequency of patient types is used to define time blocks within a day (i.e., "session"), (3) a block's time capacity is subdivided into time slots based on the patient type information, and (4) time slots within blocks are allocated to specific patient types. We acknowledge that practitioners in fact are prototyping and publishing experiments with similar scheduling policies in actual clinics (Huang and Verduzco, 2015). For example, in Huang and Verduzco (2015), the authors studied an actual women's health clinic and identified patient types and demand ratios, yet their research approaches are not very rigorous. In that work, the clinic distributes different patient types consistently over time. This example further supports that our block scheduling policy can be widely applicable for practitioners.

3.1.3 Contributions

Our essay contributes by introducing a simple and easy-to-use block scheduling policy which repeats block assignments throughout a day. This policy is grounded in practitioner developments as well as in the very successful and broadly applicable Toyota approach for manufacturing of multiple product types. The two features of our study — patient heterogeneity and patient no-shows — clearly are ones that practitioners in outpatient services are interested in, as demonstrated by recent practitioner experiments with similar block scheduling systems (Huang and Verduzco, 2015). In the U.S.A., there are over 900 million outpatient ambulatory case visits annually (CDC and Prevention, 2010), at over 96,000 outpatient care centers establishments (Business Data Codes, 2015). Thus, this essay might provide actionable managerial insights to a large number of outpatient clinic managers and schedulers.

Finally, our study can be widely applicable to other professional service organizations (e.g., financial consultations) in the context of scheduling customers of multiple types having relatively fixed service times.

To the best of our knowledge, this study is among the first few sequential scheduling studies to provide guidance for outpatient clinics. We first provide an algorithm that can lead to an optimal block schedule when considering two patient types where no-show rates are zero. Although the problem with more than three patient types is proved to be NP-Hard, meaningful insights for both practitioners and academic scholars are provided. We then consider patient no-shows along with patient heterogeneity to examine conditions under which our block scheduling policy is effective. We also adapt the block scheduling approach to implement overbooking tactics, a widely adopted policy to mitigate patient no-shows. Finally, since some healthcare providers have switched from traditional appointment scheduling policies to open-access approaches, we also consider our block scheduling policy within an open-access environment allowing same-day appointment scheduling. Due to demand uncertainty in the open-access scenario, we employ computational experiments to compare our sequential block scheduling policy against alternative open-access policies. We demonstrate how two variants of our block scheduling approach improve upon extant literature.

The remainder of this essay is organized as follows. Section 3.2 reviews related literature. Section 3.3 provides model notation and formulations for traditional scheduling situations. Section 3.4 examines overbooking with our block scheduling approach. Section 3.5 compares open-access policies adopting our block scheduling algorithms. Section 3.6 concludes and provides potential directions for future research.

3.2 Literature Review

Our research relates to the large body of appointment scheduling research in the healthcare area. A great number of operations management scholars are interested in healthcare appointment scheduling, and the topic has been explored over the last 50 years (Cayirli et al., 2012; Gupta and Denton, 2008; LaGanga and Lawrence, 2012). A critical literature survey on traditional appointment scheduling is developed by Cayirli and Veral (2003) and Gupta and Denton (2008) providing an extensive taxonomy of methodologies as well as directions for future research. In the literature, the design of appointment systems has been mainly focused on identifying appointment rule, for "within the day" scheduling that typically considers the combination of timing of patient arrivals and sequencing patients, with the objective of balancing between patient waiting time and the idle time of surgeons (Cayirli and Veral, 2003; Cayirli et al., 2012).

3.2.1 Appointment System Design

Appointment system design comprises three decisions: (i) which appointment rule to use, (ii) how to classify patients, and (iii) how to adjust scheduling for no-shows, walk-ins, and other patient events (Cayirli and Veral, 2003). An appointment rule identifies a size of scheduling time block, a length of time interval for appointment slots within a block, and initial variables such as the number of patients that will be scheduled. Pioneering stuides on appointment system design, studied extensively in the operations management field (Bailey, 1952; Ho and Lau, 1992; Soriano, 1966), explored appointment rules derived from many combinations of block size, appointment interval, and initial variables.

The second decision in appointment scheduling design considers whether to recognize patients as heterogeneous. Patient populations can be categorized using many different characteristics (e.g., new/return patient, service time differences, and arrival patterns). A decision maker can design a policy to assign patient categories to time blocks and appointment slots. This research stream suggests that adopting patient grouping approaches for new/return patients (Cox et al., 1985), inpatient/outpatient (Walter, 1973), patient care procedure types (Bosch and Dietz, 2000), and different service times (Wang, 1997) can substantially improve a clinic's schedule performance.

The third decision for designing an appointment system concerns adjustment of schedules for patient behaviors such as no-shows and walk-ins. Extant literature explores patient no-show behaviors via equal service time scheduling policies (LaGanga and Lawrence, 2012; Robinson and Chen, 2010). Empirical studies indicate patient no-shows are widespread and drive negative financial effects (Moore et al., 2001; Pesata et al., 1999). Conversely, several studies propose that refusing walk-in patients also may cause negative effects for a clinic (Cayirli and Veral, 2003; Taylor, 1984; Virji, 1990). To address issues related to patient behaviors, scheduling literature examines two types of appointment scheduling policies: traditional and open-access scheduling policies (Robinson and Chen, 2010).

3.2.2 Traditional Policies vs. Open-Access Policies

With traditional scheduling policy, a clinic sets a routine appointment schedule in advance of the day of treatment. Traditional scheduling policies do not allow same-day patient call-ins or walk-ins. Using traditional policies, a clinic can easily develop a no-idle-time schedule. However, these appointment schedules often cannot appropriately support clinic administrators due to variability in patient arrival and treatment processes (Robinson and Chen, 2010). One source of variability affecting a scheduling policy relates to patient no-shows, which are common in practice (Cayirli et al., 2006; Ho and Lau, 1992; Rust et al., 1995). If a scheduled patient does not show up, a clinic may not utilize available resources. No-shows can lead to negative financial and care quality effects (Hixon et al., 1999; Moore et al., 2001; Pesata et al., 1999). One approach to remediate no-show effects is to use an overbooking appointment policy that allows a clinic to book multiple patients in a particular time slot (Muthuraman and Lawley, 2008; White and Pike, 1964). Prior studies examine impacts of overbooking policies on patient no-shows (Feldman et al., 2014; LaGanga and Lawrence, 2012; Liu and Ziya, 2014). An overbooking policy can improve the physician's workload efficiency, but also may increase patients' waiting time.

To mitigate this trade-off, an alternative scheduling policy, called open-access scheduling, was proposed in the 1990s (Qu et al., 2007). Under an open-access scheduling environment, a clinic allows patients to call into the clinic in the morning to make a same-day appointment (Herriott, 1999; Erdogan et al., 2015; Murray and Tantau, 1999, 2000). The proportion of the schedule devoted for the daily call-in appointments is one key parameter to provision appropriate capacity to meet patient demand (Herriott, 1999; Qu et al., 2007). While some literature argues that the effectiveness of open-access scheduling is questionable (Kodjababian, 2003), these arguments are mainly based on qualitative experience.

Recent studies of open-access scheduling tend to use rigorous quantitative approaches, including stochastic optimization, queuing models, and simulation. Dobson et al. (2011) study the effect of reserved capacity for urgent patients in primary healthcare settings using a stochastic model. Their study demonstrates that when a clinic is not overloaded, the optimal scheduling policy depends on the ratio of the cost of treating an urgent patient to the cost of delaying a regular patient. Qu et al. (2007) demonstrate a quantitative method to determine the proportion of call-in/same-day time slots to derive optimized clinic appointment schedules. Green et al. (2007) investigate the impact of panel size, or the number of patients that a doctor needs to see in a day, in an open-access system. Using the overflow frequency level, which is the percentage of demand that exceeds capacity, the study provides an approach to select the optimal call-in panel size. Similarly, Green and Savin (2008) use a queuing model to estimate the impact of panel size when all patients want to make an appointment as soon as possible. Robinson and Chen (2010) analytically compare cost between traditional and open-access scheduling policies for homogeneous patients, and demonstrate that an open-access system outperforms traditional schedules unless patient waiting time is little or the probability of no-shows is minor.

Overall, our review of literature on healthcare clinic outpatient appointment systems implies that not many studies explore heterogeneous patient characteristics and patient no-shows together. Zacharias and Pinedo (2014) and Schütz and Kolisch (2013) are among the few studies that consider both patient heterogeneity and patient no-shows. Assuming equal service time for patients, Zacharias and Pinedo (2014) consider patient heterogeneity in the sense of no-show probabilities. Our work is general and is different from Zacharias and Pinedo (2014) in the sense that we allow variable service times and patient-type-dependent no-show probabilities. Schütz and Kolisch (2013) consider different patient types requiring integer multiples of time slots to explore patient heterogeneity and patient no-show issues. Our work is different from Schütz and Kolisch (2013) in the sense that we do not impose any restriction on the service time being an integer multiple of predefined time slots. Rather, the fixed time slot is defined in our essay as a function of the input parameters of the problem. In other words, the length of a time slot is a weighted average of service time requirements of different classes of customers served in a block.

In short, to the best of our knowledge, no literature provides sequential block scheduling policy for traditional, overbooking, and open-access cases, where schedules are affected by new/follow-up patients, no-shows, and call-ins. Our essay contributes by considering such factors, which often occur in the clinical environment, leading to more realistic scenarios for considering optimal scheduling policy.

3.3 Problem GP: Traditional Scheduling Policy without Overbooking

In this section, we consider Problem GP, a patient scheduling problem under the traditional scheduling environment without overbooking. Note that under the traditional system, each patient makes an appointment well in advance of the scheduled date. The clinic is able to fill up all available time slots on a given day. In Problem GP, no overbooking is allowed. Thus, only one patient is assigned to each time slot. We describe the problem setting and provide notation in Section 3.3.1. In Section 3.3.3, we consider the problem when all patients are guaranteed to show up at the appointed time. Then, in Section 3.3.4, we discuss the problem when patients' no-show probabilities are positive. Finally, in Section 3.3.5, we discuss how practitioners may want to adjust our approach into their preferred time slot set up.

3.3.1 Problem Formulation

Our research is motivated by our discussions with practitioners, practitioner-oriented literature, such as Huang and Verduzco (2015), and a pediatric orthopedic clinic case (Klassen et al., 2010). Based on the motivating literature, we design the framework for our research. We generalize the framework so the findings apply to many clinic scheduling contexts.

3.3.1.1 Motivating Clinic Case

The clinic opens daily from 8:00 AM to 12:10 PM. On average, a physician at the clinic can examine 25 patients during each one-day session. The 4 hours and 10 minutes (i.e., 250 minutes) available during a session are divided into 25 time slots with each patient taking one 10-minute time slot. The first patient is scheduled at 8:00 AM, while the last one at 12:00 PM. Overtime cost occurs if the physician continues treating patients after 12:10 PM.

Based on historical data, the clinic categorizes patients into two different types: new patients and follow-up patients. In general, a new patient has to go through more diagnostic examinations than a follow-up patient. The clinic observes that the average service time required by a new patient is 13 minutes, while the average service time required by a follow-up patient is 8 minutes. The clinic also observes that the variation in service times within each patient type is insignificant. The clinic is able to conclude that the ratio between the number of patients of the two types is very stable. In the clinic, 40 percent of patients are new patients, while the remaining 60 percent are follow-up patients. The clinic also observes that patients do not show up for scheduled appointments all the time. Also, a significant difference exists between the no-show probabilities of the two patient types.

3.3.1.2 Problem Setting

Motivated by the above clinic features, we describe a general traditional scheduling problem. A clinic provides services to two types of patients, Type A (e.g., new patients in the case clinic) and Type B (e.g., follow-up patients in the case clinic). The average service times of these two types of patients, μ_a and μ_b, are different. Without loss of generality, we assume that $\mu_a > \mu_b$. In addition, the average no-show rates of Type A patients, p_a, and Type B patients, p_b, may differ. The arrival rates of these two types of patients may also be different. The clinic is able to derive the ratio $r_a : r_b$ as the smallest positive integer ratio between the arrival rates of Type A and Type B patients.

The clinic adopts a block scheduling policy. The length of each time slot, L, is fixed and satisfies $L = \frac{r_a\mu_a + r_b\mu_b}{r_a + r_b}$. Then every $r = r_a + r_b$ time slots are considered as a block. The clinic needs to assign r_a Type A patients and r_b Type B patients in each block. Once the scheduling sequence in a block is determined, the clinic then repeats the same block sequence k times to fill up the one-day session. This scheduling approach is referred to as "sequential block scheduling policy" throughout the essay. Thus, on a given day, during the regular hours, there are $T = kr$ time slots scheduled.

We have several assumptions for the basic model: (i) The length of service time for each patient type is constant. Extant literature shows that service time variance is low in many outpatient clinics (LaGanga and Lawrence, 2012). LaGanga and Lawrence (2007) demonstrate that a scheduling policy with fixed service time is effective even in a high service time variance environment. (ii) The patients are scheduled well ahead of time. Since the demand from the patients is always higher than the capacity of a clinic, the clinic is able to fill up the time slots based on its preference. (iii) The clinic does not serve unscheduled walk-in patients (Liu and Ziya, 2014). (iv) The patients and physicians punctually arrive in a scheduled time slot, following extant literature (Klassen and Rohleder, 1996; LaGanga and Lawrence, 2012; Soriano, 1966). The following Table 3.1 introduces the notation used in our models.

Table 3.1: Notation

Parameter	**Note**
μ_a	The average service time of Type A patient. In our example above, $\mu_a = 13$ minutes.
μ_b	The average service time of Type B patient. In our example, $\mu_b = 8$ minutes.
r_a	The number of Type A patients scheduled within a block. In our example, $r_a = 2$.

Table 3.1 Continued

Parameter	Note
r_b	The number of Type B patients scheduled within a block. In our example, $r_b = 3$.
r	Total number of patients scheduled within a block, $r = r_a + r_b$. In our example, $r = 5$.
L	The length of each period (time slot). In our example, $L = 10$ minutes.
k	The number of blocks to be scheduled in the planning horizon (e.g., a day). In our example, $k = 5$.
T	The number of periods (time slots) in the planning horizon, $T = kr$. In our example, $T = 25$.
n_a	The number of Type A patients scheduled in the planning horizon.
n_b	The number of Type B patients scheduled in the planning horizon.
n	Total number of patients scheduled in the planning horizon, $n = n_a + n_b$.
p_a	The probability that a Type A patient does not show up for the scheduled appointment.
p_b	The probability that a Type B patient does not show up for the scheduled appointment.
σ_t	The type of patient assigned to time slot t, $t = 1, 2, \ldots, T$. For example, $\sigma_1 = A$, $\sigma_2 = B$.
π	The sequence of a block, $\pi = (\sigma_1, \sigma_2, \ldots, \sigma_r)$. For example, $\pi = (A, B, A, B, B)$.
s	The schedule of the planning horizon. In our example, s is defined as a concatenation of blocks: $s = (\pi, \pi, \pi, \pi, \pi)$.
α_w	Unit cost for patients' waiting time.
α_d	Unit cost for physician's idle time.
α_o	Unit cost for physician and facility overtime.
z_t	A dummy variable to indicate whether the t^{th} patient shows up, $1 \leq t \leq T$: $z_t = 1$, if the patient shows up; 0, otherwise.
$\vec{z}$	One realized instance of schedule s: $\vec{z} = (z_1, z_2, \ldots, z_n)$.
μ_t	The service time of the patient who arrives at the beginning of period t. $\mu_t = \mu_a$, if the assigned patient is Type A; $\mu_t = \mu_b$, otherwise.
b_t	The backlog at the beginning of Period t after the t^{th} patient either shows up or not, i.e., the additional amount of time to serve the first t patients at the beginning of Period t.
w_t	The waiting time of the t^{th} patient.
d_t	The physician's idle time in the t^{th} period.
$W(s, \vec{z})$	The patients' total waiting time of $\vec{z}$, a realization of schedule s.

Table 3.1 Continued

Parameter	Note
$D(s, \vec{z})$	The physician's idle time of $\vec{z}$, a realization of schedule s.
$O(s, \vec{z})$	The physician's overtime of $\vec{z}$, a realization of schedule s.
$C_w(s)$	The expected total patient waiting time cost for schedule s.
$C_d(s)$	The expected total physician idle time cost for schedule s.
$C_o(s)$	The expected total physician and facility overtime cost for schedule s.
$C(s)$	The expected total cost for schedule s, i.e., $C(s) = C_w(s) + C_d(s) + C_o(s)$.

3.3.2 *Cost Calculation*

When overbooking is not considered, the clinic assigns exactly T patients to T available time slots, i.e., $n = T$. There are two possible scenarios for each patient: she either arrives punctually or does not show up. A vector of show/no-show dummies $\vec{z} = (z_1, z_2, \ldots, z_n)$ can be used to describe a realized instance of schedule s. Also, given schedule s, we know the type of patient assigned to Period t, and consequently μ_t, the service time of the t^{th} patient, $1 \le t \le n$. Given the vector (z_t, μ_t), $1 \le t \le n$, we first describe the calculation of b_t which is used to calculate the costs.

Calculation of Backlog b_t: The backlog, b_t, is defined as the additional amount of time to serve the first t patients at the beginning of Period t. Thus, when $t = 1$, we have $b_1 = z_1\mu_1$. For $t \ge 2$, there are two possible scenarios: (1) If $b_{t-1} \le L$, then the physician can finish serving the first $(t-1)$ patients before the start of Period t. Thus, $b_t = z_t\mu_t$; (2) If $b_{t-1} > L$, then the physician cannot finish serving the first $(t-1)$ patients before the start of Period t. The extra time needed is $(b_{t-1} - L)$. In this scenario, b_t has the value $b_{t-1} - L + z_t\mu_t$. Thus, to combine these two scenarios, we have:

$$b_t = \begin{cases} 0, & \text{if } t = 0, \\ (b_{t-1} - L)^+ + z_t \mu_t, & \text{if } t \geq 1. \end{cases}$$

Calculation of Expected Total Cost $C(s)$: Three types of costs are considered as follows:

- Expected total patient waiting time cost ($C_w(s)$)

  Given $\vec{z}$, a realized instance of schedule s, to calculate the waiting time cost, we first derive the waiting time of the t^{th} patient, w_t. There are three possible scenarios: (i) If a scheduled patient does not show up, then her waiting time is considered as 0. (ii) The waiting time of the patient who is scheduled for the first period is 0, since there are no patients scheduled before her. (iii) For any patient who shows up at the beginning of Period t ($t \geq 2$), her waiting time depends on b_{t-1}. If $b_{t-1} \leq L$, then the t^{th} patient does not need to wait. Otherwise, the patient needs to wait an additional $(b_{t-1} - L)$ minutes to get her service. Thus, we have:

  $$w_t = \begin{cases} 0, & \text{if } t = 1 \text{ or } z_t = 0 \\ (b_{t-1} - L)^+, & \text{if } t \geq 2 \text{ and } z_t = 1. \end{cases}$$

  Accordingly, the patients' total waiting time for $\vec{z}$, a realized instance of schedule s, is $W(s, \vec{z}) = \sum_{t=1}^{n} w_t$. Thus, the patient waiting time cost is:

  $$\alpha_w W(s, \vec{z}) = \alpha_w \sum_{t=1}^{n} w_t.$$

  The above calculation is for $\vec{z} = (z_1, z_2, \ldots, z_n)$, one realized instance of schedule s. Since each of the n patients can either show up or not, there are in total 2^n scenarios. The probability associated with each scenario is $p(\vec{z}) =$

$(1-p_a)^{x_a} p_a^{r_a - x_a} (1-p_b)^{x_b} p_b^{r_b - x_b}$, where x_a (resp. x_b) represents the number of Type A (resp. Type B) patients who show up in instance $\vec{z}$. Thus, the expected total waiting time cost is:

$$C_w(s) = \sum_{\vec{z}} p(\vec{z}) \alpha_w W(s, \vec{z}).$$

- Expected total physician idle time cost ($C_d(s)$)

  At the beginning of Period t, after the t^{th} patient either shows up or not, we obtain b_t. If $b_t < L$, then a physician will be idle after spending b_t minutes with the patient in Period t. Thus, the physician's idle time in Period t is $d_t = L - b_t$. If $b_t \geq L$, then the physician will be treating the patient during Period t. Thus, $d_t = 0$. Therefore, we have $d_t = (L - b_t)^+$. The physician's total idle time of one instance $\vec{z}$ of schedule s is $D(s, \vec{z}) = \sum_{t=1}^{n} d_t$. Thus, the physician idle time cost for $\vec{z}$, a realization of schedule s, is:

$$\alpha_d D(s, \vec{z}) = \alpha_d \sum_{t=1}^{n} d_t.$$

  Thus, the expected total waiting time cost of schedule s is:

$$C_w(s) = \sum_{\vec{z}} p(\vec{z}) \alpha_d D(s, \vec{z}).$$

- Expected total physician overtime cost ($C_o(s)$)

  A physician's overtime depends on b_n. When $b_n > L$, the physician needs extra time to care for the remaining patients. Thus, we have $O(s, \vec{z}) = (b_n - L)^+$. Thus, the physician overtime cost for $\vec{z}$, a realization of schedule s, is:

$$\alpha_o O(s, \vec{z}) = \alpha_o (b_n - L)^+.$$

Thus, the corresponding expected total overtime cost of schedule s is:

$$C_w(s) = \sum_{\vec{z}} p(\vec{z})\alpha_o O(s, \vec{z}).$$

Combining together, the above cost components, the expected total cost for schedule s is:

$$C(s) = C_w(s) + C_d(s) + C_o(s).$$

Thus, the objective is to find an optimal schedule s^* such that $C(s^*) = \min_s C(s)$.

3.3.3 When Patients' No-Show Probabilities Are Zero

The results developed in this section are used later to develop scheduling approaches for traditional scheduling environments without overbooking (Section 3.3) and with overbooking (Section 3.4). The results also are used in developing open-access scheduling policies later in Section 5, as having zero no-show probabilities is a reasonable assumption in this environment.

In Problem GP, the objective is to minimize the expected total cost ($C(s)$) of schedule s. However, previous studies suggest that physician-related per-unit time cost is much higher than patient per-unit waiting time cost (Cayirli et al., 2012; Zacharias and Pinedo, 2014). Due to this reason, a clinic may prefer a class of schedules that minimizes the physician's idle time and overtime. Thus, we introduce Problem SP, a variant focusing on schedules that minimize a physician's expected idle time and overtime. In this section, we study Problem SP_0, a variant of Problem SP in which the no-show probabilities of both types of patients are 0, i.e., $p_a = p_b = 0$. Thus, for each schedule s, there is only one possible instance, i.e., $z_t = 1$, $t = 1, 2, \ldots, n$. In Problem SP_0, we only consider the schedules with zero physician idle time and overtime. Among the schedules with zero physician idle time and overtime, we try to select one which minimizes the patients' waiting time.

$$\text{Problem } SP_0\text{:} \quad \min_{s} \quad C_w(s)$$

$$\text{s.t.} \quad C_d(s) = 0,\ C_o(s) = 0.$$

In Section 3.3.3.1, we first describe Set Λ, which represents the feasible set of Problem SP_0, i.e., the class of block schedules having zero physician idle time and overtime. We also provide an upper bound for Problem SP_0. Section 3.3.3.2 introduces an algorithm to derive an optimal solution of Problem SP_0. Finally, Section 3.3.3.3 discusses computational complexity of the problem with $m \geq 3$ patient types.

3.3.3.1 Set Λ: Feasible Set of Problem SP_0

Let Set Λ represent the set of schedules that satisfy the following properties: (i) The schedule is a block schedule, (ii) Each patient's no-show rate is zero, (iii) The schedule has zero physician idle time, and (iv) The schedule has zero physician overtime. Note that in each block, r_a Type A patients and r_b Type B patients are assigned to $r = r_a + r_b$ periods. The length of the block equals the sum of the service times of these r patients, i.e., $rL = \sum_{i=1}^{r} \mu_i$. In the following lemma, we describe the conditions for a block schedule to have zero physician idle time and overtime. The reader may refer to the online supplement for all proofs of Lemmas and Theorems.

Lemma 1 *A block schedule π has zero physician idle time and overtime if and only if $\sum_{i=1}^{t} \mu_i \geq tL$, $\forall 1 \leq t \leq r$.*

In the online supplement, we analyze the upper bound of patients' waiting time for any schedule in Set Λ.

3.3.3.2 Algorithm OptBlock(π_o) to Derive an Optimal Solution of Problem SP_0

We now study the schedule in Set Λ which minimizes patients' waiting time. We propose Algorithm *OptBlock(π_o)*, an algorithm to derive π_o. Thus, $s_o = (\pi_{o1}, \pi_{o2}, \ldots, \pi_{oi}, \ldots, \pi_{ok})$ is an optimal solution for SP_0.

Algorithm *OptBlock*(π_o)

Begin

Set $J_a = r_a$, $J_b = r_b$.

Assign Type A patient to Period 1 of π. Set $F_1 = \mu_a$, $J_a = J_a - 1$, and $t = 2$.

While $(J_a + J_b > 0)$ do

Step 0: If $F_{t-1} + \mu_b \geq tL$ and $J_b > 0$, then perform Step 1;

otherwise, perform Step 2.

Step 1: Assign Type B patient to Period t of π.

Set $F_j = F_{j-1} + \mu_b$, $J_b = J_b - 1$, $t = t + 1$.

Step 2: Assign Type A patient to Period t of π.

Set $F_j = F_{j-1} + \mu_a$, $J_a = J_a - 1$, $t = t + 1$.

End(while)

Output: π_o

End

We briefly explain the idea behind Algorithm *OptBlock(π_o)*. Note that $\mu_a > L > \mu_b$. To guarantee there is no physician idle time in the first time slot, we must assign it to the type of patient with a longer service time (Type A patient in our example). Next, whenever possible, the scheduling policy tries to assign the patient with shorter service time (Type B patient in our example), as long as doing so will not incur any idle time. We define the sequence obtained by this algorithm as "*OptBlock* Sequence."

Example of OptBlock Sequence Schedule: For the example parameter settings specified in Section 3.3.1.2, Algorithm *OptBlock(π_o)* produces "ABABB" as the *OptBlock* Sequence, and accordingly, the schedule for a whole session of 25 time slots would be $|ABABB||ABABB||ABABB||ABABB||ABABB|$.

Since the while loop performs r_a+r_b iterations, we know that our scheduling policy can be performed within polynomial time. Thus, the computational complexity of Algorithm *OptBlock(π_o)* is $O(r_a + r_b)$. Lemma 2 states this complexity. We show optimality of *OptBlock* Sequence based policies in Lemma 3.

Lemma 2 *The complexity of Algorithm OptBlock(π_o) is $O(r_a + r_b)$.*

Lemma 3 *Algorithm OptBlock(π_o) provides an optimal schedule for Problem SP_0.*

In summary, our policy offers an optimal schedule within set Λ. In total, this policy maximizes labor resource utilization and minimizes patients' waiting time.

3.3.3.3 Generalization to m Patient Types with Zero No-Show Probability

Although most outpatient clinics categorize patients into two types, some clinics may want to consider more general cases of patient heterogeneity. This section extends patient heterogeneity to $m \geq 3$ general types. We first define Problem SP_0^m as follows.

Problem SP_0^m: Given there are m ($m \geq 3$) types of patients with different service time, find a block schedule s_o with zero physician idle time and zero overtime such that the total patients' waiting time W_{s_o} is minimized.

We first describe Algorithm $OptBlock_m(\pi)$, a variant of our Algorithm $OptBlock(\pi_o)$ that is adapted for m types of patients. Algorithm $OptBlock_m(\pi)$ finds schedule $s_o = (\pi, \pi, \ldots, \pi)$, where π is a block sequence consisting of ($r = r_1 + r_2 + \ldots + r_m$) patients. We name the m types of patients as Type i, $i = 1, 2, \ldots, m$. Without loss of generality, we assume $\mu_1 < \mu_2 < \ldots < \mu_m$.

Algorithm $OptBlock_m(\pi)$

Begin

Set $J_l = r_l$, $l = 1, 2, \ldots, m$. Set $F_0 = 0$ and $t = 1$.

While ($\sum_{l=1}^{m} J_l > 0$) do

Step 1: Find the smallest index i ($1 \leq i \leq m$)

such that $F_{t-1} + \mu_i \geq tL$ and $J_i > 0$.

Step 2: Assign Type i patient to Period t of π.

Set $F_t = F_{t-1} + \mu_i$, $J_i = J_i - 1$, and $t = t + 1$.

End(while)

Output: π

End

In the following observation, we demonstrate that Algorithm $OptBlock_m(\pi)$ may not derive an optimal solution to scheduling problem SP_0^m for fixed $m \geq 3$.

Observation 1: Algorithm $OptBlock_m(\pi)$ may not provide an optimal solution for fixed $m \geq 3$.

Although Algorithm $OptBlock_m(\pi)$ may not provide an optimal solution, it may provide a good heuristic solution. We summarize the two feasibility conditions for any solution π: (i) The physician's idle time is 0 in each period. (ii) There are exactly r_i Type i patients assigned in π^*, $i = 1, 2, \ldots, m$.

Next, we describe DP_m, a dynamic programming (DP) algorithm which can be used to derive π^*, a block sequence to form an optimal solution of Problem SP_0^m. We define $\vec{\sigma}_t = \{\sigma_1, \sigma_2, \ldots, \sigma_t\}$ as the vector that includes the types of patients assigned from Period 1 to Period t, and define $J_i(\vec{\sigma}_t)$ as the number of Type i patients that have been assigned in $\vec{\sigma}_t$. At the beginning of Period t, $1 \leq t \leq r$, for each feasible $\vec{\sigma}_{t-1}$, we first obtain the set which includes all possible indexes i ($1 \leq i \leq m$) such that $\sum_{l=1}^{t-1} \mu_{\sigma_l} + \mu_i \geq tL$ and $J_i(\vec{\sigma}_{t-1}) < r_i$. Then we create $\vec{\sigma}_t$ by including one index from the feasible set at a time. In the next lemma, we describe the computational complexity of DP_m.

Lemma 4 *The complexity of DP_m is $O(m^r)$.*

Note that for fixed r and m, m^r is polynomial in m. Thus we have the following result.

Lemma 5 *DP_m solves Problem SP_0^m polynomially for fixed r and m.*

If r and m are variables (which means they are parts of the input), then problem SP_0^m is a strongly NP-hard problem.

Theorem 1 *Problem SP_0^m is strongly NP-hard when r and m are variables.*

Although many clinics classify patients into two types, some clinics categorize their patients into more than three types. Theorem 1 provides an important message for outpatient clinic managers: categorizing patients into three or more types will increase the difficulty to obtain an optimal schedule.

3.3.4 When Patients' No-Show Probabilities Are Positive

We next study Problem GP when patients may not show up. Note that if patients do not show up, a clinic experiences physician idle time. In Section 3.3.4.1, we show in Theorems 2 and 3 that schedules in Set Λ outperform other schedules regarding the physician's expected idle time and overtime. Then in Section 3.3.4.2, we use numerical experiments to compare the performances of schedules in Set Λ and other schedules when patients' expected waiting time cost is also considered.

3.3.4.1 Two Properties of Schedules in Set Λ

In this section, we analyze the performance of schedules in Set Λ in Problem GP. In Theorem 2, we first compare among schedules in Set Λ.

Theorem 2 *For any schedule in Set Λ, the expected physician idle time has the same value.*

Next, in Theorem 3, we compare any schedule in Set Λ with any schedule that is not in Set Λ.

Theorem 3 *For any schedule that is not in Set Λ, the physician's expected idle time and overtime is larger than that of any schedule in Set Λ.*

Recall that Problem SP is a variant of Problem GP focusing on schedules minimizing the physician's expected idle time and overtime. Thus, Theorem 3 shows that any schedule in Set Λ is an optimal solution of Problem SP. In other words, regardless of the patient no-show rate, the clinic should always consider the schedules in Set Λ if its focus is to minimize the physician's expected idle time and overtime.

3.3.4.2 Numerical Experiment to Examine the Optimality of OptBlock Sequence

As extant literature shows (Zacharias and Pinedo, 2014), an optimal schedule can vary depending upon patient no-show probabilities. Although we mathematically prove that *OptBlock* Sequence leads to an optimal scheduling solution under zero no-show probability, this sequence may not always be optimal under positive no-show probabilities. Thus, we now explore the impact of no-show probabilities p_a and p_b on *OptBlock* Sequence. We standardize the coefficient of physician's idle time as 1. Following extant literature (Cayirli et al., 2012), we assume the ratio between physician costs is $\alpha_o = 1.5\ \alpha_d$. We then examine different values of patient waiting time (α_w) from 0.1 to 0.8 as in Zacharias and Pinedo (2014). We expect that physician overtime or physician idle time is more costly than patient waiting time. Thus, schedules with positive overtime may perform worse than our schedule.

We use the following experiment to verify the performance of our scheduling policy in positive no-show cases. Table 3.2 provides numerical parameters for the

Table 3.2: Parameters for the experiment

Case	μ_a	μ_b	r_a	r_b	*OptBlock* Sequence	Range of p_a and p_b	k	L	α_w
1	22	14	1	7	$ABBBBBBB$	0.1, 0.2, ..., 0.8	4	15	0.1, 0.3, 0.5, 0.8
2	24	12	2	6	$ABBBABBB$				
3	30	6	3	5	$ABABBABB$				
4	22	8	4	4	$ABABABAB$				
5	18	10	5	3	$AABAABAB$				
6	17	9	6	2	$AAABAAAB$				
7	16	8	7	1	$AAAAAAAB$				

experiment. As described in the Problem Formulation (Section 3.3.1), we keep a block size r fixed at 8 time slots, which accounts for eight patients within a block. Then, we vary the number of Type A patients (r_a) and Type B patients (r_b) from 1 to 7. Thus, the combination of r_a and r_b can generate 7 different cases. We can obtain the *OptBlock* Sequence in each scenario. The experiment covered a range of patient no-show probabilities from 10% to 80%. Thus, one scenario can obtain 64 different cases (8 possible values of p_a, and 8 possible values of p_b). We fix the length of each time slot L as 15 minutes and fix the length of the session as 480 minutes, which is 8 regular hours per session. Thus, the experiment includes 32 time slots within a session. The corresponding number of blocks k is a determined value based on other parameters, such as μ, r, and L. Using different α_w, we calculate the total expected cost (C) to compare *OptBlock* Sequence schedules against other schedules. With the above parameter settings, we have $1,792$ different cases (i.e., 7 clinic scenarios x 64 no-show cases x 4 weights of patient waiting time). We also examined schedules that are not in set Λ and the results show that schedules in set Λ provide better performance than other schedules in most cases.

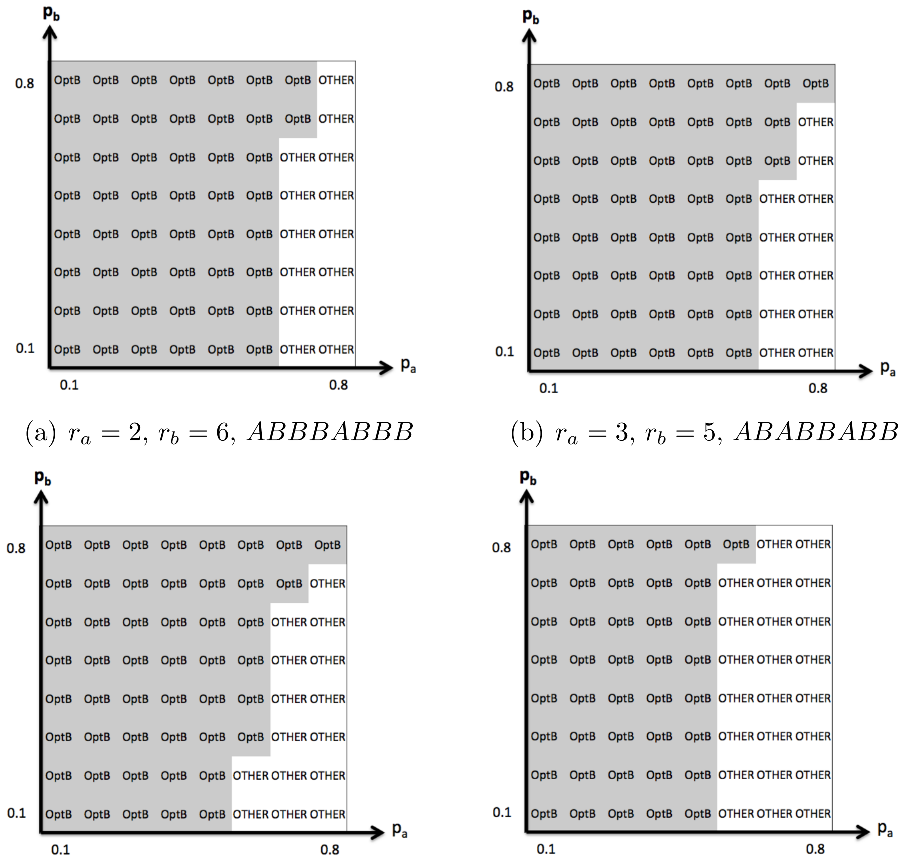
To examine the effect of our scheduling policy relative to other schedules, we first compared schedules within Set Λ, since in Theorem 3, we show that the schedules in Set Λ minimizes the physician's idle time and overtime for any range of no-show rate. For example, there exist seven feasible schedules in Λ when $r_a = 3$ and $r_b = 5$. In Table 3.2, the "*OptBlock* Sequence" column represents sequences developed by Algorithm *OptBlock*. For each p_a value, we change each patient no-show probability p_b to investigate the impact of no-shows on *OptBlock* Sequence.

3.3.4.3 Managerial Insights

Following accepted simulation approach (Law, 2013) programmed an experiment in the *C++* language, we randomly generate the number of Type A and Type B patients to schedule for a session, and then use a binomial distribution for each patient type to generate no-shows. With a fixed session length (i.e., 480 minutes), it is possible to randomly generate too many patients for a session, thus excess patients are scheduled on the following day. Thus, we need to consider the warm-up period for consistent estimation. After generating 1,100 session replications, we dropped the first 100 results, to obtain a total of 1,000 session replications for each scenario, representing approximately three years of daily sessions. We used analysis of variance (ANOVA) to examine whether the schedules of the *OptBlock* Sequence and other schedules in the set Λ (i.e., in Figure 3.2, OptB vs. OTHER) were significantly different. In each case, patient waiting time cost is statistically different between the two categories (i.e., OptB vs. OTHER).

Figure 3.1 illustrates four instances (i.e., Case No. 2, 3, 5, and 6) of experiment results, which show the total expected cost (C), as we vary patient no-show probability. The "OptB" (i.e., grey) region indicates an optimal area using the *OptBlock* Sequence and the "OTHER" (i.e., white) region indicates an optimal area using schedules other than *OptBlock* Sequence. This numerical experiment highlights implications of *OptBlock* Sequence. First, the result remains consistent across

Figure 3.1: *OptBlock* sequence cost efficiency



(a) $r_a = 2$, $r_b = 6$, $ABBBABBB$

(b) $r_a = 3$, $r_b = 5$, $ABABBABB$

(c) $r_a = 5$, $r_b = 3$, $AABAABAB$

(d) $r_a = 6$, $r_b = 2$, $AAABAAAB$

all α_w patient waiting time cost cases. This observation suggests that Algorithm *OptBlock(π_o)* can be widely applicable in many different clinic settings. Second, as shown in Figure 3.1, we observe that *OptBlock* Sequence performance is associated with Type A patients, who require more service time than Type B patients. If the no-show rate in Type A patients is not high (e.g, less than 50% in our cases), *OptBlock* Sequence clearly outperforms (i.e., minimum C) regardless of the no-show rate in Type B patients. This observation provides an important insight for clinic schedulers. If a clinic can reduce the no-show rate of patients with longer service time, the clinic can obtain the best schedule with *OptBlock* Sequence. Although our numerical experiments are not exhaustive, the experimental results demonstrate that our scheduling policy is effective in different r_a, r_b, μ_a, and μ_b environments. This insight further illuminates that level scheduling policy can be widely applicable in outpatient clinics.

3.3.5 Alternative Scheduling Policy Based on OptBlock Sequence

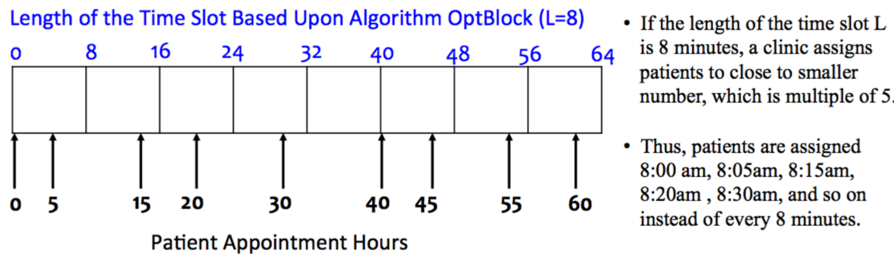
Our numerical experiment findings so far provide evidence that our policy generates effective schedules for two patient types. We now discuss how our approach might be relaxed by practitioners to adapt it to specific time scheduling preferences. In practice, schedulers may prefer to use time slot lengths close to an integer number, often a multiple of 5 or 10 minutes. That is, the preferred scheduled appointment times may be 8:00 am, 8:05 am, 8:15 am, 8:20 am, 8:30 am, 8:40 am, 8:45 am, 9:00 am, and so on. Figure 3.2 illustrates this alternative scheduling adaptation that will generate schedules closer to the desired scheduling environment.

Suppose in a clinic, we have the following parameter values: $r_a = 2$, $r_b = 3$, $\mu_a = 14$, $\mu_b = 4$. The corresponding length of block, $r_a\mu_a + r_b\mu_b$, is 40 minutes, and the length of each time slot, L, is 8 minutes. Theoretically, a clinic can assign the i^{th} patient scheduled at time $(i-1)L$ with our block scheduling policy. That is for $i = 1, 2, 3, 4, 5, 6$, the ideal scheduled appointment time should be 0, $L = 8$, $2L = 16$, $3L = 24$, $4L = 32$, $5L = 40$, $6L = 48$, $7L = 56$, $8L = 64$. However, the scheduler instead can assign patients to arrive at an earlier preferred time based on 5 minute increments (Figure 3.2).

3.3.5.1 Main Difference

To calculate cost for each session, we need to consider a variable-interval L_i. We define L_i as the actual length of the time slot for the ith patient. Thus, functions of

Figure 3.2: Practical scheduling based upon algorithm *OptBlock(π_o)*



b_t, w_t, d_t, and O must be revised. In other words,

$$b_t = \begin{cases} 0, & \text{if } t = 0, \\ (b_{t-1} - L_{t-1})^+ + z_t\mu_t, & \text{if } t \geq 1. \end{cases}$$

$$w_t = \begin{cases} 0, & \text{if } t = 1 \text{ or } z_t = 0 \\ (b_{t-1} - L_{t-1})^+, & \text{if } t \geq 2 \text{ and } z_t = 1. \end{cases}$$

$$d_t = (L_t - b_t)^+.$$

$$O = (b_n - L_n)^+.$$

The practical implication of this approach is that solutions based on *OptBlock* Sequence should have the same physician idle time and overtime, whereas the patient waiting time will be extended slightly.

In summary, we have identified that *OptBlock* Sequence can perform well for positive patient no-show rates. Overall, the *OptBlock* Sequence based schedule provides a lower expected total cost C than other schedules when the no-show rate of the patients with longer service time is not too high. In situations where patient no-show rates are high, extant works (Chen and Robinson, 2014; Robinson and Chen, 2010) have suggested the use of overbooking policy to mediate patient no-show behaviors.

Thus, we next study the impact of *OptBlock* Sequence when applied to overbooking scenarios.

3.4 Problem OP: Traditional Scheduling Policy with Overbooking

In the previous section, we study the scenario when patients may not show up for appointments. The physicians experience idle time if no overbooking is allowed. In practice, to improve the utilization rate of physicians and facilities, clinics may consider overbooking, i.e., assign more than one patient for each time slot. In this section, we study a traditional scheduling policy combining *OptBlock* Sequence with overbooking. For the overbooking model, we assume that (i) If a time slot is overbooked, then patients assigned to this time slot are of the same type (i.e., all are either Type A or Type B), (ii) There is no priority between the overbooked patients. If all patients scheduled for the same time slot arrive, the clinic will randomly select one of these patients to serve first. After serving this patient, the clinic will continue randomly selecting one patient among those arrived at the same time slot to serve. The clinic finishes serving all these patients before serving the patients scheduled after this time slot. (iii) The clinic must serve all scheduled patients even when a regular session is over. Thus, the clinic may remain open after TL minutes have elapsed.

3.4.1 The Design of Overbooking Policies

An overbooking policy handling multiple patient types should include the following three elements: (a) the type of patient assigned to each time slot of the schedule, (b) the number of overbooked patients for each type, and (c) the allocation of these overbooked patients. In this section, we propose two overbooking policies: (i) Front Load Overbooking (Policy FLO) and (ii) Level Load Overbooking (Policy LLO). We first describe the three elements of each policy, then briefly explain the rationale behind our choices. Some illustrative examples are presented to derive insightful observations.

- Type of patient assigned to each time slot of the schedule

  In both Policy FLO and Policy LLO, we use *OptBlock* Sequence as the base schedule, i.e., to determine the type of patient to be assigned to each time slot. In our example, the base schedule is

  $|ABABB||ABABB||ABABB||ABABB||ABABB|$.

  Our Rationale: In the previous section, we have shown that the schedule based on *OptBlock* Sequence is the optimal policy when patient no-show rates are zero. Thus, when no-show rates are positive and overbooking is allowed, if only those overbooked patients do not show up, then this schedule minimizes the patients' waiting time while achieving zero physician idle time and overtime. We have also shown that this policy minimizes the expected physician idle time and overtime, when patients have positive no-show rates and overbooking is not allowed. Thus, if the no-show patients spread out across the schedule, we still believe our proposed overbooking schedules below will perform well.

- Number of overbooked patients for each type

  We use Type A patients as an example. Recall that kr_a represents the number of time slots reserved for Type A patients in the base schedule. We define E_a as the number of overbooked Type A patients in the schedule. Thus, the expected number of arrived Type A patients is $(kr_a + E_a)(1 - p_a)$. Since the no-show rate p_a is positive, to minimize the physician's idle time, we would prefer the expected number of arrived Type A patients is close to kr_a. Thus, we obtain $E_a = \left\lceil \frac{kr_a}{1-p_a} \right\rceil - kr_a$. We can use a similar method to calculate the number of overbooked Type B patients. As an example, let us suppose we have

$r_a = 2, r_b = 3, k = 5$. Let $p_a = 0.15$ and $p_b = 0.25$. Thus, $E_a = 2, E_b = 5$. We need to overbook 2 Type A patients and 5 Type B patients.

- Allocation of these overbooked patients

  The idea of front loading has been used for single-type patient overbooking (LaGanga and Lawrence, 2012; Zacharias and Pinedo, 2014). By assigning all overbooked patients in the first time slot, the clinic is able to minimize the expected physician's idle time. Since we are dealing with two types of patients in our schedule, we first describe the allocation plan for our Policy FLO:

  Policy FLO

  Step 1: Identify the first time slot assigned for a Type A (resp. Type B) patient.

  Step 2: Allocate E_a (resp. E_b) Type A (resp. Type B) patients in this time slot.

  FLO Schedule S_F: $|A(2)B(5)ABB||ABABB||ABABB||ABABB||ABABB|$

  $A(2)$ means two Type A patients are overbooked to the first time slot. $B(5)$ means five Type B patients are overbooked to the second time slot.

  Next, we describe the allocation plan for Type A patients in our Policy LLO (The allocation plan for Type B patients follows a similar method):

Policy LLO

Algorithm for Policy LLO

Begin

Set $J_a = E_a$. Set $i = 1$.

While $(J_a > 0)$ do

Step 1: Find the first period k in Block i such that the patient assigned to period k is Type A and period k has not been overbooked yet.

Step 2: Overbook one Type A patient to period k of Block i.

Step 3: Set $J_a = J_a - 1$. If $i = r$, then $i = 1$; otherwise $i = i + 1$.

End(while)

Output: s

End

LLO Schedule S_L:

$|A(1)B(1)ABB||A(1)B(1)ABB||AB(1)ABB||AB(1)ABB||AB(1)ABB|$.

Next, to compare the performances of our two overbooking policies, we study several scenarios under which E_a Type A and E_b Type B scheduled patients did not show up.

- Scenario OB1: The last E_a Type A and the last E_b Type B scheduled patients did not show up.
- Scenario OB2: The first E_a Type A and the first E_b Type B scheduled patients did not show up.
- Scenario OB3: The E_a Type A and E_b Type B patients who did not show up are evenly distributed.

We have the following observations:

Observation 2: Under Scenario OB1, schedule S_L (LLO) outperforms schedule S_F (FLO).

Under both S_L and S_F, we overbook E_a Type A patients and E_b Type B patients. If the last E_a Type A and the last E_b Type B scheduled patients did not show up, then the system utilization rate remains 100% in both schedules. However, in S_F, patient waiting time is much higher than that in S_L due to the high number of patients assigned on the front.

Observation 3: Under Scenario 0B2, schedule S_F (FLO) is the optimal schedule. The physician's idle time and overtime are zero. The patient waiting time is minimized.

When the first E_a Type A and the first E_b Type B scheduled patients did not show up in a session, the resulting schedule becomes a traditional appointment schedule without overbooking: $|ABABB||ABABB|$ $|ABABB||ABABB||ABABB|$. We have proved the optimality of this policy in the previous section.

As a consequence of the above observations, we have the following result.

Observation 4: Under all three scenarios, schedule S_F (FLO) has zero idle time.

Remark: Scenario 3 is the likely occurrence in practice. Under Scenario OB3, it is difficult to conclude which one of the two schedules (S_L, S_F) dominates. We expect that S_L (FLO) dominates S_F (LLO) for the most practically relevant problem instances defined in our computational study. We explore this in our computational study in the next section.

3.4.2 Numerical Experiment for Overbooking

We adopt the experiment setting used for the numerical experiments in Section 3.3.4.2 for our simulation, fixing number of patients per block, $r = 8$, and overall session hours, TL, which are equivalent to 480 minutes. In addition, we include three

Table 3.3: Parameters for overbooking experiment

$Case$	μ_a	μ_b	r_a	r_b	Sequence	Range of p_a and p_b	k	L	α_w
8	40	8	3	5	$ABABBABB$	0.1,0.2,...,0.7,0.8	3	20	0.1, 0.3, 0.5, 0.8
9	30	10	4	4	$ABABABAB$				
10	26	10	5	3	$AABAABAB$				

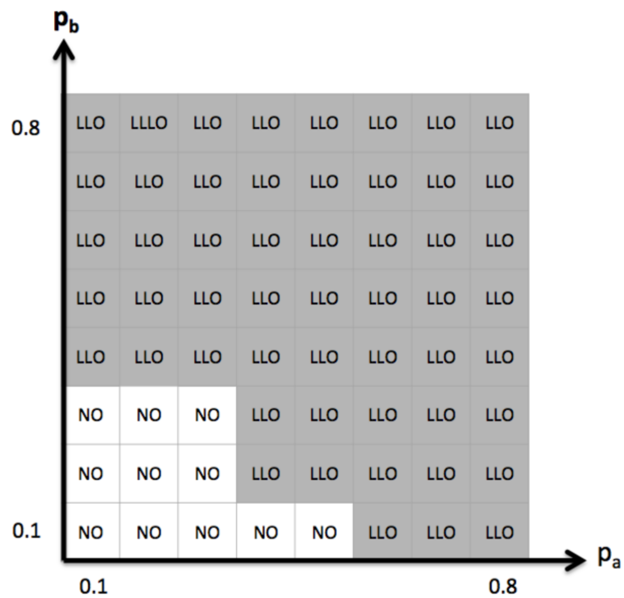
more cases to examine the impact of L on scheduling performance. Table 3.3 shows the additional cases on top of the previous cases in Table 3.2. Finally, we estimate the effect of no-show probabilities on the optimal overbooking policy when we control for other parameters. We then compare scheduling without overbooking (NO) and with overbooking policies (LLO and FLO). We first examined the analysis of variance (ANOVA) to ensure all three policies (i.e., LLO vs. FLO vs. NO) are significantly different.

3.4.2.1 Managerial Insights

In general, we observe that the LLO policy always provides better performance (i.e., lowest C) than the FLO policy in terms of marginal total cost that includes patient waiting time cost, physician idle time cost, and physician overtime cost. This result further supports the idea that Toyota's production leveling philosophy can lead to efficient service system utilization and effective care delivery in overbooking settings. Since LLO provides a better overbooking policy, we now examine the impact of overbooking policy using LLO scheduling.

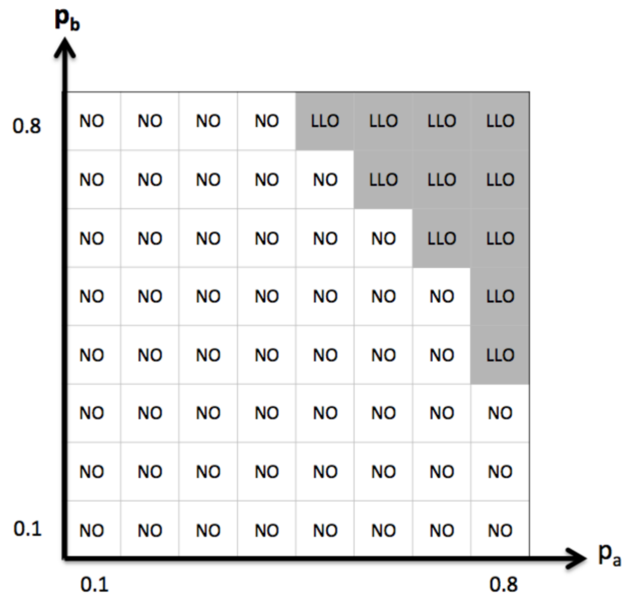
Figure 3.3 and Figure 3.4 illustrate one instance (i.e., Case No. 3 in Table 3.2) of simulation results that compare overbooking with leveling scheduling (LLO) against the no-overbooking policy (NO). When the total cost (C) per patient for LLO is lower than NO, we label the cell *"LLO"*, otherwise we label it *"NO"*. The X axis

Figure 3.3: LLO vs.NO overbooking ($\alpha_w = 0.1$)



and Y axis indicate the no-show rate of Type A and Type B patients, respectively. Figure 3.3 provides the result when the coefficient of patient waiting time is 0.1. Figure 3.4 shows results when the patient waiting time coefficient is 0.5. Overall, *LLO* performs well under the high patient no-show rates. Another observation is the following: as α_w increases, the region of *LLO* shrinks. Other scenarios, which have different r_a, r_b, μ_a, and μ_b parameters, follow similar patterns. One implication from the simulation is as follows: a clinic, which has physician idle time that is 10 times more expensive (i.e., $\alpha_w = 0.1$) than patient waiting time, will be better off to overbook patients if no-show rate is moderate. If a clinic considers patient waiting time to be at least half the value of physician idle time, then it is better off not to overbook patients unless the patient no-show rates are extremely high.

Figure 3.4: LLO vs.NO overbooking ($\alpha_w = 0.5$)



3.4.2.2 Regression Analysis

Since our numerical results cannot estimate distinct clinic characteristics that can affect the benefits of overbooking, we now consider a regression analysis to examine effects of patient characteristics on a clinic's decision to overbook. Specifically, we use a logistic regression model to predict whether characteristics of patient demand, no-show, and service length can encourage clinics to use *LLO* overbooking. Using *DOverbook* as a dependent dummy variable, we estimate the logistic regression model: $Logit(DOverbook) = \beta_0 + \beta_1 Service + \beta_2 Demand + \beta_3 p_a + \beta_4 p_b + \beta_5 L + e$, where β_0 is constant and e is a random error term. The dependent variable (*DOverbook*) takes the value 1 when an "*LLO*" policy is better to use than "*NO*", and 0 otherwise. We use service time ratio (*Service*)(i.e., $\frac{\mu_a}{\mu_b}$) and demand ratio (*Demand*)(i.e., $\frac{r_a}{r_b}$) as determinants of benefits of overbooking. We also include control variables (p_a, p_b, and L).

Table 3.4: Determinants of overbooking decision

Variable	$\alpha_w = 0.1$	$\alpha_w = 0.5$	$\alpha_w = 0.8$
Service	-0.243	-0.427***	-0.357***
	(0.18)	(0.127)	(0.13)
Demand	-0.027	0.001	0.047
	(0.11)	(0.07)	(0.07)
p_a	17.879***	11.116***	10.148***
	(2.10)	(0.92)	(0.89)
p_b	15.286***	10.242***	8.566***
	(1.80)	(0.87)	(0.79)
L	-0.076	0.026	0.041
	(0.07)	(0.05)	(0.05)
cons	-6.21***	-11.03***	-10.915***
	(1.33)	(1.09)	(1.10)
N	768	768	768
R^2	0.682	0.554	0.504

$^{*}\ p < 0.1$ $^{**}\ p < 0.5$ $^{***}\ p < 0.01$

Table 3.4 provides estimated coefficients. The regression results demonstrate that the probability of benefits from overbooking a schedule decreases as the service time ratio between Type A and Type B increases (*Service*). However, if $\alpha_w = 0.1$, the service time ratio is not significant. In other words, the impact of the service time ratio matters most when a clinic places a high value on a patient's time, relative to the value placed on the clinic's idle time (i.e., $\alpha_w > 0.5$). Next, there is no evidence that the probability of benefits from overbooking a schedule change as the demand ratio between Type A and Type B patients changes (*Demand*). Also, the probability of benefits from overbooking a schedule increases as the patient no-show rate (p_a, p_b) increases. Finally, we observe no empirical evidence of effects of the length of time slot (L) on the overbooking benefit.

In summary, the simulation results with *OptBlock* Sequence overbooking demonstrate that overbooking, when using a level scheduling approach, can mitigate effects of patient no-shows. The results show the effectiveness of level scheduling for overbooking policy. As previously described, the use of open-access scheduling policy is the other way to mediate impacts of patient behaviors. Thus, we next study the impact of *OptBlock* Sequence on the open-access policy.

3.5 Open-Access Scheduling Policy

To mitigate the impact of patient no-shows, some clinics implement an open-access scheduling policy, where patients are allowed to make same day appointments, rather than an overbooking policy. Robinson and Chen (2010) have demonstrated the relative dominance of open-access schedules for homogeneous patients under zero patient no-show rates, as compared against traditional and overbooking schedules experiencing typical no-show rates. Motivated by this implication, we consider whether scheduling policies based on *OptBlock* Sequence perform well within the constraints of an open-access scheduling environment. We explore whether open-access schedules generated by policies based on *OptBlock* Sequence show better performance than an extant open-access scheduling policy. A key idea built into an open-access scheduling policy is the allowance for same-day appointments, in order to benefit from their observed low or zero no-show rates. In this section, we introduce two scheduling policies based upon *OptBlock* Sequence to adapt into the open-access scenario. We then implement simulation experiments to evaluate their performance. We take the base case scheduling scenario for this simulation from previous literature (Robinson and Chen, 2010), which adopts the First-Come, First-Appoint (FCFA) rule. Since environmental assumptions in an open-access schedule differ from the assumptions in Section 3.3, we start by introducing our problem statement.

3.5.1 Problem Statement

In academic open-access literature, clinics generally require patients to call early in the morning on the same day before starting their operations (Chen and Robinson, 2014; Robinson and Chen, 2010). Since literature demonstrates that the observed no-show rate in open-access scheduling environments can be significantly lower than that in traditional scheduling (Cayirli et al., 2012; Chen and Robinson, 2014), we also assume that the no-show rate in the open-access policy is negligible. Therefore, following this literature, we assume there is zero no-show probability for both patient types.

For justification of an open-access clinic's regular hours and maximum extended working hours, we adopt an empirical report, which surveyed American physicians (Hawkins, 2012). The survey results support the parameter settings used in our simulation models. We assume that a clinic's regular working hours are 8 hours (Cayirli et al., 2006).

We also assume that an open-access clinic does not allow walk-in patients or call-in patients during regular operating hours. In other words, a clinic will only consider patients who actually call in the morning, in line with the assumptions of Robinson and Chen (2010). The arrivals of the two types of patients are independent and follow separate Poisson distributions, a reasonable assumption based upon scheduling literature (Cayirli et al., 2012). The additional parameters are the following.

Additional Parameters:

λ_a	The daily demand rate of Type A patients.
λ_b	The daily demand rate of Type B patients.

3.5.1.1 *Challenge of Implementing Open Access Policy for Multiple Patient Types*

With traditional scheduling, a clinic knows the exact number of patients of each type to schedule in a given session. In contrast, under an open-access policy, the clinic needs to assign the patient to a time slot when she calls in the morning. At this point, the clinic scheduler can only observe the patients who have already called in, but has no idea how many more patients will call in later that morning. This feature is less of an obstacle when all patients are assumed to be of the same type. The clinic can just assign the patients consecutively into the slots available. However, when there are multiple types of patients, the situation becomes more complicated. As we have shown in Section 3.3.4, to balance waiting cost against idle time and overtime costs, it is optimal to assign different types of patients to different time slots by using the sequential block schedule. However, in an open-access situation, the uncertainty of the actual demand makes an optimal schedule unlikely. Thus, we apply our block scheduling approach in two different ways. Below we describe the three approaches we compare:

- First Come First Appoint (FCFA) Policy – (Benchmark Policy)
  The clinic scheduler assigns patients to the next available slot regardless of their types. The shortcoming of this policy is that the schedule in each block may not be optimal, given that the patients call in randomly. For example, we may have patients call in the following order: $BBBBA$. By using the FCFA policy, we schedule these five patients in this same order. However, if the optimal block schedule is $ABABB$, the physician likely will have a positive idle time under this policy.

- Strict *OptBlock* Policy

  Our first open-access approach is based on the *OptBlock* Sequence in Section 3.3.4. Thus, a patient will only be assigned to a slot reserved for the same type. The shortcoming of this policy is we may have too many partially filled blocks at the end of a day. For example, on a day with a large excess of Type A patients and no excess of Type B patients, any additional blocks opened to satisfy excess demand would exhibit physician idle time. Thus, the physician idle time may make this schedule inefficient in an open-access setting.

- Flexible *OptBlock* Policy

  We also propose a Flexible *OptBlock* policy, which resolves the possible shortcomings of the above two policies to some extent. In this approach, we first use the *OptBlock* policy to assign the first r_a Type A patients (resp. r_b Type B patients) to the slots based on their type. But, if we receive a call from the $(r_a + 1)^{th}$ Type A patient before filling the current block, instead of opening a new block, we temporarily switch to the FCFA policy to schedule. Then, this Type A patient is assigned to the first reserved slot for a Type B patient in this unfilled block. For example, if patients call in the following order: $BBBAB$, then using the Strict *OptBlock* policy for the first four patients, we have $AB?BB$ with the third slot unfilled. Since the fifth patient is still of Type B, we then use FCFA to assign this patient to the 3rd slot to obtain the schedule: $ABBBB$. In this instance, the schedule obtained will be better than the schedule obtained by using either of the first two policies.

Among the above three policies, we expect none will always be better than the other two. By conducting a thorough experiment, we show that either the Strict *OptBlock* policy or the Flexible *OptBlock* policy is the most robust one. Next, we provide an algorithm which describes the Flexible *OptBlock* policy used to obtain a schedule for the open-access environment.

Algorithm $Flexible(s)$

Begin

Step 0: Set $J = 1$, $n_a = 0$, $n_b = 0$.

While (A new patient calls in) do

Begin

Step 1: If the new patient is of Type A, then $n_a = n_a + 1$, go to Step 2;
otherwise, $n_b = n_b + 1$,go to Step 3.

Step 2: If $n_a \leq r_a$, assign this Type A patient
to the first available slot in Block J which is reserved for Type A;
otherwise, assign this Type A patient to the first available slot in
Block J which is reserved for Type B. Go to Step 4.

Step 3: If $n_b \leq r_b$, assign this Type B patient to the first available slot
in Block J which is reserved for Type B; otherwise, assign
this Type B patient to the first available slot in Block J
which is reserved for Type A. Go to Step 4.

Step 4: If $n_a + n_b = r_a + r_b$, set $K = K + 1$, $n_a = 0$, $n_b = 0$.

End

Output: s

End

3.5.1.2 Modified Overtime Cost Calculation for Open-Access Scheduling

To calculate cost for each open-access session in open-access scheduling, we introduce additional parameters. Now t_k is defined as the actual time slot for the kth patient. Next, t_f is the time slot that the last patient is assigned. Since we assume that a clinic can obtain specific demand information about Type A and Type B patients, the clinic scheduler knows the approximate number of needed blocks. In this open-access scenario, the calculations of expected physician idle time cost (C_d) and expected patient waiting cost (C_w) are the same as in the basic model cost calculation. Thus, we only describe the new formulation of expected physician overtime cost (C_O). Since we define T as the number of periods (time slots) in the planning horizon (a day) without overtime, we can derive the physician overtime as: $O = (b_{t_f} + L(t_f - 1) - LT)^+$.

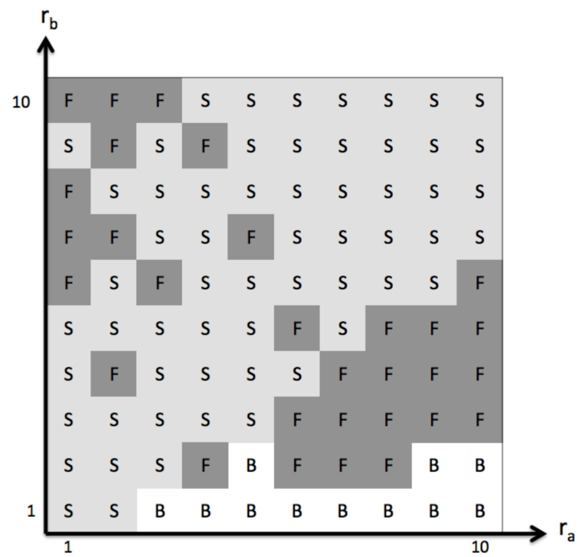
3.5.2 Performance Evaluation of Open-Access Scheduling Policies

We perform simulation experiments to compare three scheduling policies (i.e., FCFA Policy, Strict *OptBlock* Policy, and Flexible *OptBlock* Policy). Again, the base scheduling policy is the FCFA policy that prior work (Robinson and Chen, 2010) adopted for open-access.

As we mentioned, the daily number of patient early morning call-ins follows a Poisson distribution. A clinic will experience a volume of calls from Type A patients with call-in ratio λ_a ($= kr_a$), and from Type B patients with λ_b ($= kr_b$) call-in ratio. Recent open-access scheduling literature (Chen and Robinson, 2014; Robinson and Chen, 2010) justifies the credibility of this assumption.

We first examine the impact of patient numbers within a block (i.e., r_a and r_b) on the performance of scheduling policies based on *OptBlock* Sequence. We adopt fixed service time parameters for Type A patients as 20 minutes, and Type B patients as 10 minutes. These expected service times are acceptable in actual practice (Hawkins, 2012). Then, we vary the number of Type A patients (r_a) and Type B patients (r_b) from 1 to 10. Thus, the combination of r_a and r_b generate 100 different cases. Since we fix r_a, r_b, and demand ratios as integer numbers, we cannot make the exact same

Figure 3.5: Open-access total cost

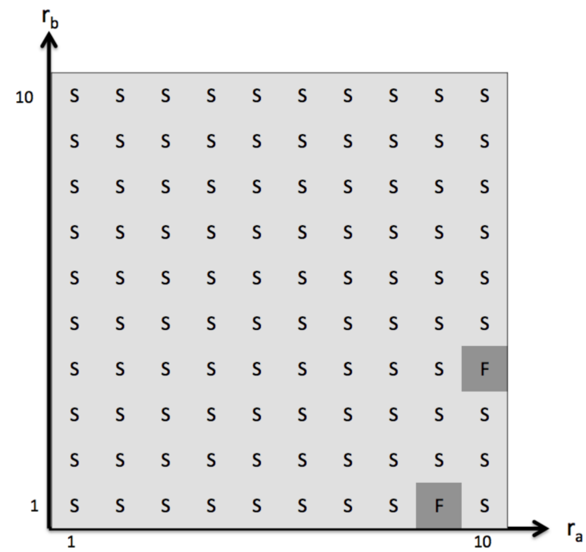


amount of the length of session (TL) in each case. Instead, we try to make each case such that the length of session is close to 240 minutes (i.e., one session as a half day). To calculate the total cost (C), we use the same parameters for cost coefficients that we adopt in Section 3.4 (i.e., α_d= 1, α_o= 1.5, and α_w = 0.1, 0.3, 0.5, and 0.8).

Figure 3.5 illustrates the results. In the figure, F indicates the Flexible *OptBlock* policy, S indicates the Strict *OptBlock* policy, and B indicates the Benchmark FCFA policy. Overall, schedules based on *OptBlock* Sequence perform well for the open-access scheduling environments. The results support that scheduling policies based on *OptBlock* Sequence can improve system efficiency when $\frac{r_a}{r_b} < 2.5$. One implication identified through the result is the following: a patient category, which requires more service time, may play a significant role in open-access scheduling.

The finding also supports the effectiveness of scheduling policies based on *OptBlock* Sequence by minimizing patient waiting time. Specifically, Figure 3.6 shows that the Strict *OptBlock* Policy often provides better schedules that have lower ex-

Figure 3.6: Open-access patient waiting time

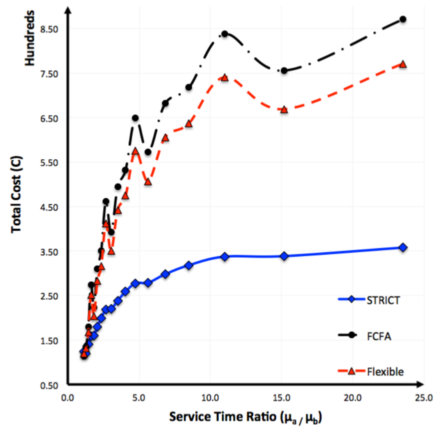


pected patient waiting time than the base policy. Also, the Strict *OptBlock* Policy performs better than the Flexible *OptBlock* Policy for most of the environments. Therefore, the combined results for the impact of patient numbers within a block supports the efficiency and effectiveness of *OptBlock* Sequence in open-access situations.

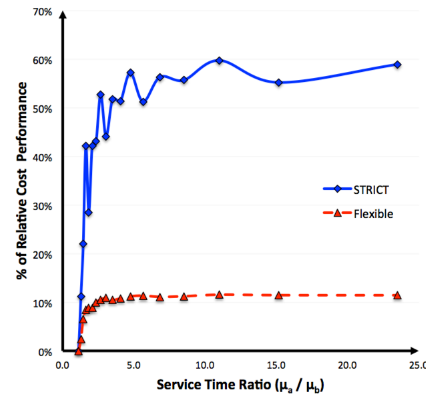
We next examine the impact of service time ratio (i.e., $\frac{\mu_a}{\mu_b}$) on the performance of scheduling policies based on *OptBlock* Sequence. While the previous experiment examines the relationship between patient numbers and performance of *OptBlock* Sequence, the patient service time ratio ($\frac{\mu_a}{\mu_b}$) is fixed. We next vary the service time ratio, while keeping the length of a regular session constant. We also vary the length of time slot (L = 10, 15, 20, and 30) and the number of patients within a block ((r_a, r_b) = (2,3) and (3, 2)). We also use the same parameters for cost coefficients to calculate the total cost (C).

Figure 3.7 illustrates results of open-access scheduling policies in service time

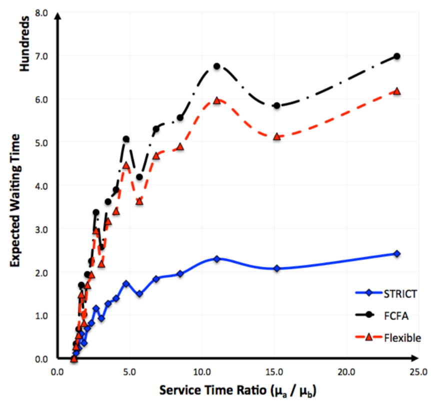
Figure 3.7: Open–access results in service time ($\frac{\mu_a}{\mu_b}$)



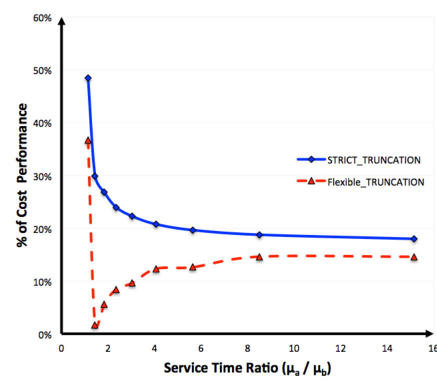
(a) Total Cost vs. Service Time Ratio



(b) Relative Cost Performance vs. Service Time Ratio



(c) Patient Waiting Time vs. Service Time Ratio



(d) Relative Cost Performance Between Truncated vs. Base Scenarios

ratio ($\frac{\mu_a}{\mu_b}$). Specifically, Figure 3.7(a) shows the total cost (C), as we vary the service time ratio, for the vector (r_a, r_b) = (2,3) and L=10. A dominant policy shown in the figure is the Strict *OptBlock* Policy. The result indicates that the total cost increases as the service time ratio increases, and as the service time ratio increases, "F" and "S" are better than "B". Figure 3.7(b) presents relative cost performance of "F" and "S" against the benchmark scheduling policy (i.e., FCFA). If the service time ratio is higher than 1.3, then *OptBlock* Sequence based scheduling policies perform dominantly in terms of the total cost C. When we vary parameters for the sensitivity analysis, "F" and "S" perform better than the "B" policy. Further, the experiment still supports the effectiveness of *OptBlock* Sequence in various service time ratios. Figure 3.7(c) shows the expected patient waiting time against service time ratio, where again "F" and "S" perform better than the "B" policy.

Above, we assumed that a clinic accepts all patients who call in during the early morning. However, in practice, many clinics may limit the size of the patient panel during a session, depending upon the clinic capacity. Thus, we modify the simulation design so that a clinic will turn down patients who call-in after the $(\lambda + 1)^{th}$ patient. In other words, the scheduling system allows patients up to λ_a for Type A and λ_b for Type B patients. Figure 3.7(d) demonstrates that *OptBlock* Sequence based scheduling policies are even more efficient in the truncated open-access environment. Thus, the experimental results suggest scheduling policies based on *OptBlock* Sequence can be adopted efficiently in actual clinics.

We have identified that *OptBlock* Sequence can perform well in situations where a clinic will use an open-access scheduling policy. Either the Strict or Flexible schedule developed by *OptBlock* Sequence provides a lower total cost C than the base scheduling policy (FCFA). Thus, we argue that the benefits of scheduling policies based on *OptBlock* Sequence still hold when applied to open-access environments.

3.6 Conclusion

We examine an outpatient appointment scheduling system under patient heterogeneity and patient no-shows. Our research extends prior healthcare operations research on scheduling to policies for patient heterogeneity. Our sequential block scheduling approach is generated from the idea of production leveling used in the Toyota Production System. Specifically, we adapt an approach for leveling of production requirements so that product mix and volume are relatively even over time. The objective of production planning aims to balance the workload in each work station. To the best of our knowledge, our work is the first research to apply this idea to healthcare scheduling systems.

This article has several implications. First, when considering two heterogeneous patient types, we develop a sequential block scheduling policy that leads to efficient and effective appointment schedules. Since this policy is easily implementable, outpatient clinics, which distinctly face patient heterogeneity, should benefit from adopting our scheduling policy to schedule patients. Second, if the clinic faces more than three patient types, clinic schedulers are encouraged to use the proposed dynamic programming procedure to find the base block schedule. Third, although our policy may not provide optimal schedules incorporated with positive patient no-shows, our policy nevertheless performs well when the no-show rate for patients who require longer service time is not significantly high. Fourth, using a logistic regression, we identify causal factors for outpatient clinic managers to consider when examining whether to use an overbooking policy. Finally, we demonstrate the impact of our scheduling policies on open-access environments, which allow same-day appointments. Since extant literature (Robinson and Chen, 2010) shows that an open-access approach can perform better than traditional and overbooking scheduling policies, and we showed our block scheduling approach performed better than FCFA in the open-access environment, we argue that our scheduling policy should provide better schedules in many outpatient scheduling settings. The findings contribute to the outpatient scheduling literature by bridging the research gap and providing stepping stones for future scheduling research.

We observe many fruitful opportunities for future research. Prior research compares open-access scheduling having zero no-show rates against traditional overbooking scheduling with positive no-show rates. Future studies of multiple patient types may relax this assumption and explore the impact of our block scheduling policy. Researchers also can further investigate impacts of constraining schedules having appointment times to multiples of 5 or 10 minute increments. Finally, we foresee many research opportunities for detailed overbooking analysis. Thus, future research can enhance our block scheduling approach within even more varied overbooking environments.

4. EMPIRICAL ANALYSIS OF HOSPITAL BEHAVIORS RESULTING FROM HEALTHCARE FINANCIAL INCENTIVE POLICY

4.1 Introduction

A long-standing concern facing the U.S. healthcare system pertains to low care delivery quality and the existence of medical errors associated with a lack of effective service delivery by healthcare providers (Naveh et al., 2005; Green, 2012). Given the complexity of healthcare service delivery, parties including policy makers, medical professionals, and academic researchers argue that minimizing process variation is one of the key drivers to reduce medical errors (Schmenner, 2004; Tucker, 2004; Tucker et al., 2007). In past years, U.S. medical errors have led to the deaths of about 100,000 patients annually with more than \$3 billion of unnecessary additional costs (Adamy, 2014; Kohn et al., 1999). To address these issues and improve objective service quality and the perceived quality of healthcare service delivery, the U.S. Medicare program has implemented several incentive and penalty systems that require hospitals to consider and demonstrate their service delivery effectiveness. Among them, Medicare in 2011 revamped its reimbursement system to develop the Value-Based Purchasing (VBP) program, which forces care providers (i.e., hospitals) and care professionals (i.e., doctors) to become responsible for improving healthcare service quality (Werner and Dudley, 2012).

As part of the VBP program, Medicare can withhold a certain amount of reimbursements (1% in 2013, 1.25% in 2014, and later up to 2%) from hospitals that do not perform well along a specified list of healthcare quality outcome metrics (CMS, 2014b). In contrast, hospitals that perform exceptionally well can receive incentive bonuses. Through the measurement of hospital care processes, patient satisfaction, and care outcomes, hospitals that participate in the VBP program are assigned either penalties or incentives. Among the nearly 3,000 hospitals that were required to participate in VBP, the program penalized about 1500 hospitals in 2013 and 2014, based on hospital operating data from the previous year and two years before the performance period, calculated as weighted scores pertaining to poor patient satisfaction and low process quality. Hospitals incurred total financial penalties of about $1.1 billion (CMS, 2014b).

Since the VBP penalty eventually will not be negligible for most hospitals, given the annual growth of the penalty and uncertain hospital profit margins, there is an expectation for hospital healthcare managers to comply with this government regulation. Nevertheless, instead of triggering a hospital to make significant process and care outcome quality improvements, some hospitals may make opportunistic adjustments to avoid penalties. Although potentially illegal to do so, in response to VBP, hospitals may not be willing to admit high-risk patients (i.e., risky health conditions, low-income patients, or certain races of patients) who are likely to be conducive to poor quality healthcare performance (Jha et al., 2010). Hospitals also may respond to VBP penalties by tactically focusing on new patient sectors that provide stronger remuneration. Healthcare professional skepticism about the effects of the VBP program on process quality improvement (Rau, 2013) supports the need for careful, detailed scrutiny of the VBP program and related initiatives.

We investigate the VBP program to examine its quantitative impacts on healthcare quality improvements when hospital managers face this institutional pressure. Prior studies demonstrate healthcare organizations are concerned with unavoidable external pressures (Scott et al., 2000; Lee and Zenios, 2012; Ata et al., 2013). To respond to this VBP institutional pressure, some hospitals may adopt symbolic management practices to comply with social standards in appearance (Westphal and

Zajac, 1994), leading to symbolic practices that may not align with VBP program expectations. Organizational theory scholars have explored the symbolic practice phenomenon, which often occurs when an organization needs to demonstrate external legitimacy, but cannot afford to acquire substantial resources to modify practices (Fiss and Zajac, 2006; Westphal and Graebner, 2010). Healthcare policy makers and hospital administrators, as well as U.S. citizens, need to know how much of a real effect the VBP program actually has on healthcare providers and the overall healthcare system. Thus, we raise the following unexplored research question: *Are financially penalized hospitals likely to adopt tactics relevant to symbolic management practices in response to penalties from VBP?*

To address our research question, we employ several different data sets from the Center for Medicare & Medicaid Services (CMS), which provides detailed hospital data including process quality, patient experience, and other related information. We also obtain specific VBP measures and hospital environment data from the Hospital Compare data from Medicare. Using information from HIMSS Analytics, we obtain hospital information technology adoption data. Finally, we obtain information from the Dartmouth Atlas, which provides information on geographical distribution of hospitals.

Our research endeavors to contribute both to academic research and practitioners in the healthcare sector. First, compared to emerging OM studies that hypothesize the VBP program will uniformly improve hospital operational performance, this study theorizes contextual differences in the behaviors of hospitals when facing these external government pressures. By empirically examining evidence of drivers of symbolic practice, our research introduces the symbolic perspective into healthcare service OM research. Next, little empirical research in healthcare OM estimates responses to the variety of exogenous institutional pressures intended to improve healthcare processes and outcomes, thus this essay contributes empirical evidence by quantifying impacts of the VBP program, while controlling for other relevant regulatory programs. As Green (2012) suggests, managing patient-oriented service processes is an essential topic for the future of the OM field. Finally, since the VBP program is a touchstone financial program, among several programs intended to motivate the objectives of patient-centered care, our empirical analysis of the VBP program contributes to key principles that the OM field should endeavor to move toward: disentangling actual operating improvements from symbolic improvements motivated by financial incentives.

For government policy makers and hospital administrators, we estimate the contemporary effects of the current VBP program. We investigate whether the program pushes hospitals to opt to use symbolic practice instead of process improvement practices. Our findings suggest that when hospitals are financially penalized by the VBP program, then the hospitals are more likely to respond in appearance. Specifically, previously penalized hospitals may avoid more complicated patients while accommodating patients who can bring extra revenues. Through considering these findings, government policy makers might be motivated to consider additional metrics that may generate more effective incentives for improving the overall healthcare system. Otherwise, the VBP program may lead to unexpected negative associations with the intended healthcare quality improvements. In addition, there are nearly 3,000 VBP participating hospitals in the U.S.A. that annually discharge about 10,000 patients per hospital on average (CMS, 2015a). Thus, the findings might provide actionable managerial insights directly related to the care quality and outcomes for 30,000,000 patients per year affected by the VBP program. Given our findings, if government policy makers can empirically identify that a hospital exhibits symbolic practice outcomes, then hospital administrators will need to carefully control the stakeholder behaviors to counteract unintended, or potentially illegal, operational processes. If a hospital is exhibiting symbolic practices in the short-run, administrators might need to initiate organizational changes for the long-run to ensure compliance with the VBP intentions.

In Section 4.2, we provide an overview of the VBP program, review literature on institutional theory and symbolic management perspectives, and generate hypotheses. In Section 4.3, we describe data and methods. We present results in Section 4.4. Finally, we discuss implications for both practitioners and researchers about the empirical findings.

4.2 Background and Hypotheses

We first review aspects of government regulations in the healthcare industry. We then explore theoretical perspectives underlying organizational actions based on institutional symbolism. We pose hypotheses based on symbolic management literature pertinent to healthcare provider reactions.

4.2.1 The Value-Based Purchasing (VBP) Program

Government-initiated programs in healthcare have made efforts to motivate improved quality of medical service. These programs require healthcare providers to meet certain rules, standards, and expectations. U.S. healthcare-related government policy, regulatory bodies, and reimbursement bodies, such as Medicare and the Department of Health and Human Services (HHS), have developed sets of best practice protocols, which provide a general approach for process management practices for care process measures (Benner and Tushman, 2003; Chandrasekaran et al., 2012). These process management practices, inspired by Total Quality and Lean Practice adoption, support healthcare providers in the quest to deliver consistent services and procedures (Westphal et al., 1997). Corresponding to recommendations of the Institute of Medicine (IOM) to tackle the quality of healthcare service, the Center for Medicare & Medicaid Services (CMS), for instance, in 2003 developed quality measurement programs for reducing process variation and promoting quality improvements (Boyer et al., 2012; Kohn et al., 1999). CMS provided best-practice process measures for hospitals to assess general and severe health issues, spanning from best practices for the timing of antibiotic treatment for general surgery patients to best practices for the quick response to heart attack patients. Healthcare providers who participate in Medicare or Medicaid programs must provide related data to CMS to verify they are conforming to such practices.

In 2010, CMS released a new regulation called the Value-Based Purchasing (VBP) program, which connects the Medicare payment system to care delivery quality metrics. The program's purpose is to reduce cost and to improve healthcare quality (Rau, 2012). To do so, Medicare can charge reimbursement penalties or provide reimbursement bonuses based on a hospital's annual quality measures and actual healthcare outcomes during prior years (CMS, 2014b). Table 4.1 provides the set of metrics for the VBP payment program. Regarding quality measures, the underlying rationale for the VBP program is the consideration of two sets of performance qualities and one set of care outcomes. The first quality measure concerns process of care adherence to internal clinical procedures in which a healthcare provider follows the CMS recommended guidelines when they treat patients. For instance, for each of the four conditions measured (e.g., heart failure), one of the VBP clinical measures is the percentage of hospital patients whose antibiotic selection was appropriate. The second measure relates to perceived patient experience, which is an external quality measure that considers how care providers deliver services to patients. For example, the degree of nurse or doctor communications with patients is one external quality measurement dimension. While the two sets of measures estimate process quality of

Table 4.1: Key metrics for the VBP payment program

	Description of Measures	Note	2013	2014	2015
Process of Care Measures	• Acute Myocardial Infarction (AMI) • Heart Failure (HF) • Pneumonia (PN) • Surgical Care Improvement Project (SCIP)	CMS measures each metric's performance and improvement rates	✓	✓	✓
Patient Experiences	• Communication • Responsiveness • Pain Management • Hospital Environment Conditions • After Discharge Satisfaction	Hospital Consumer Assessment of Healthcare Providers and Systems Survey (HCAHPS) measures patient satisfaction using patient survey data	✓	✓	✓
Outcome of Care Measures	• Heart Failure • Heart Attack (AMI) • Pneumonia	Mortality rate of patients who died within 30 days after being treated for these conditions		✓	✓

the care providers' internal care delivery protocols and patient satisfaction, the VBP program also evaluates the care providers based on a third set of objective outcome of care measures, such as mortality of three health conditions (Heart Failure, Acute Myocardial Infarction (AMI), and Pneumonia), each of which is viewed as a crucial disease that results in high rates of death and hospitalization, and accordingly, excessive costs for all stakeholders (i.e., patients, care providers, third party payers, and U.S. taxpayers).

Previous management literature already suggests government regulations and policies are associated with healthcare executive actions (Elsbach et al., 1998; Oliver, 1991; Pfeffer, 1981; Ruef and Scott, 1998). Goodrick and Salancik (1996) and Oliver (1991) argue that depending on the characteristics of care providers, the degree of compliance with a regulation will differ, and the constraints under which top management's strategic choice is made should be investigated. Some studies examine the impact of healthcare management on quality (Chandrasekaran et al., 2012; Shortell et al., 1995). However, very little empirical research tackles specific behavior of healthcare providers when they address institutional pressures such as government-initiated VBP quality improvement programs intended for healthcare providers. Recent working papers simply analyze the VBP-to-care quality linkage. We contribute by demonstrating how hospitals can respond to VBP either through actual process improvement or instead via symbolic practice.

4.2.2 Institutional Pressures and Symbolic Management

Institutional theorists argue that external pressures, such as government regulations, will influence the motivation of organizational behaviors, making firms behave differently than would be expected under strictly logical and rational actions (DiMaggio and Powell, 1983; Scott et al., 2000). Institutionalization refers to societal procedures by which external policies obtain legitimacy in an organization (Meyer and Rowan, 1977; Westphal et al., 1997). Indeed, organizational literature proposes legitimacy as "a generalized perception or assumption that the actions of an entity are desirable, proper, or appropriate within some socially constructed system of norms, values, beliefs, and definitions" (Suchman, 1995, p. 574). To acquire legitimacy, organizations are more likely to operate using similar strategies or decision-making systems, a situation called an isomorphism (DiMaggio and Powell, 1983; Heugens and Lander, 2009).

Based on the institutional perspective, there exist three types of isomorphic processes: coercive, mimetic, and normative processes (DiMaggio and Powell, 1983). Coercive isomorphism comes from formal or informal pressures exerted on organizations by governments or other dominant agencies upon which they depend. Prior literature relates to coercive isomorphism exerted by government regulators (Barratt and Choi, 2007), customers (Choi and Eboch, 1998), and headquarters (Kostova and Roth, 2002). Mimetic isomorphism refers to the cognitive isomorphic process in which organizations recognize institutionalization as taken-for-granted beliefs

(DiMaggio and Powell, 1983; Scott et al., 2000). Organizations are likely to face unexpected environments that entail risks. Through a mimetic isomorphism process, the organizations follow best practices within an industry to tackle economic peril and acquire legitimacy (Heugens and Lander, 2009). Finally, the normative isomorphic process refers to institutionalization as a pursued value among peers of a professional network, such as professional organizations, trade associations, and public opinion (DiMaggio and Powell, 1983; Scott et al., 2000). These groups may impose pressures on organizations to conform to specific standards (Peng, 2003). In response to the three isomorphic processes, organizations may enact different strategic responses including symbolic practice (Markóczy et al., 2013; Oliver, 1991).

Symbolic management literature suggests organizations under institutional pressures instead are likely to adopt symbolic practices conforming to social expectation (i.e., legitimacy) in appearance, without conforming actual operating practices (Westphal and Zajac, 1994, 1998, 2001; Zajac and Westphal, 1995). We define symbolic practice as the adoption of organizational legitimacy in name only, where internal organizational structures move in an opportunistic direction. Organizational theorists empirically demonstrate that the organizational symbolic management perspective is positively associated with instances of high implementation cost, or high symbolic gain (Elsbach et al., 1998; Martinez-Moyano et al., 2013; Rogers et al., 2007; Westphal and Graebner, 2010). For example, hospitals are likely to exert symbolic practices when they want to diminish patient attention, such as with highly charged issues of hospital billing practices (Elsbach et al., 1998). In healthcare OM literature, symbolic practice has been explored using case-based research (Bhakoo and Choi, 2013). Using the information technology adoption context, the study explores whether hospital symbolic practices vary depending on institutional pressure characteristics.

In summary, the institutional and symbolic management literatures suggest explanations for the ways institutionalization occurs and for drivers of institutional pressures and symbolic practices. Yet most prior studies use a qualitative approach to examine such issues in healthcare. Little research empirically studies drivers of symbolic management or the impact of institutional pressures in the healthcare industry. We contribute by investigating the symbolic practice phenomenon in the context of VBP within the healthcare industry via an econometric approach.

4.2.3 Research Hypotheses

4.2.3.1 Prior Performance and Government-Regulations

Government mandates and corresponding symbolic management literature suggest that prior performance in terms of poor hospital performance threatens the reliability of hospitals as an appropriate service provider for patients (Scott et al., 2000; Westphal and Zajac, 1994). To assuage this quandary, hospital administrators should at least provide some signal to patients (including potential patients) that the hospital conforms to the particular government mandate performance criteria, at least in appearance (Oliver, 1991; Pfeffer, 1981; Scott et al., 2000). The following year's hospital performance represents the hospital managers' efforts to improve service quality in a way that satisfies patients (DiMaggio and Powell, 1983). While hospitals can use many forms of tactics for symbolic practices (Fiss and Zajac, 2006), the operational performance of the following year is the most direct and externally visible outcome. The VBP program's disclosure of performance metrics for a hospital is observable to patients and provides hospitals with a signaling mechanism demonstrating apparent service quality capability. This VBP evaluation should be based on a hospital's actual performance. However, the observed performance result may not accurately reflect on the actual hospital care delivery quality improvement efforts.

To conform to legitimacy in appearance, without substantive quality improvement, some hospitals may opportunistically manipulate patient inputs or process outputs. Assuming little or no change in exogenous patient needs in the local population (i.e., city), one way to make opportunistic adjustments relates to the change in a hospital's patient distribution. Changing the patient case mix index (CMI) is one tactic hospital managers might use to improve the hospital's quality score without making real process improvements. The case mix index refers to a hospital's level of clinical complexity for inpatient services (CMS, 2015a). Thus, a change in this index may represent a signal to indicate a change of the patient distribution. For example, if CMI increases, a hospital is treating a patient population consisting of proportionally more complicated patients.

Also, hospitals might choose to serve more patients who can bring in extra financial revenues to the hospitals. If a hospital admits more than a certain portion of a specific patient group, the hospital can achieve extra financial incentives. In doing so, rather than improving processes, they may simply attempt to counteract the impact of the VBP penalty. The disproportionate share hospital (DSH) percentage indicates the proportion of these patient groups. Further, the outlier payments percentage, measuring the proportion of patients who require extra costs to treat, represents another index based upon which a hospital receives additional financial incentives from the government. In response to the VBP financial penalty, hospitals may accept more patients who are eligible for this outlier payment category.

Motivated by symbolic institutional theory, we conjecture that when hospitals are penalized in the previous period, they are more likely to avoid patients who require more complex treatments and to report more patient proportions that can bring in incentive based financial bonuses. In short, through adjusting patient groups, specifically those groups that are negatively associated with the VBP quality measures (e.g., more complicated patients), hospitals may obtain a lower CMI compared to the previous year. Also, hospitals can also accept more patients who are directly associated with the financial incentives by increasing the proportion of DSH patients and Outlier Payment patients. Thus, symbolic managerial behavior can be reflected by a decreasing CMI, increasing DSH, or increasing Outlier Payment level. Through tactical handling of these patient groups, hospitals may avoid financial penalties from CMS, which represents conformity to VBP objectives in appearance but not in action, that is, symbolic practice. Taken together, we posit the following hypothesis:

Hypothesis 1 (H1). *The lower a hospitals previous performance, the higher the likelihood that the hospital exhibits symbolic management practices.*

4.2.3.2 Impact of Dense Referent Group on Symbolic Behavior

Since institutionalization concerns adaptive responses that are logically connected by organizational regulative, mimetic, and cognitive characteristics (DiMaggio and Powell, 1983), the diffusion of the VBP program into hospitals can be conventionally explained by such institutionalization. In the context of a hospital's referent group, which represents a source of organizations having similar organizational characteristics, different yet socially similar hospitals may take part in VBP and develop similar normative beliefs (Festinger, 1954; Reichers, 1985). These common beliefs may precede similar practice adoption among the hospitals. Researchers in management argue that organizations positioned in different referent groups engage in distinguishable strategies (Fiss and Zajac, 2004; Lounsbury, 2007). Organizational behavior within a geography base is one way to classify hospitals into different social normative systems. Studies in the institutional literature have empirically identified that organizations may take various responses to institutional pressures depending on location (Doshi et al., 2013; Lounsbury, 2001, 2007; Marquis, 2003; Marquis et al., 2007). Marquis (2003) illustrates that the density of intra-organizational networks differs across U.S. cities, and these geographical differences can lead to the adoption of different behaviors.

We conjecture that referent group density of institutional constituents enables hospitals to induce symbolic practices. When hospitals are located within a city (or county) referent group, the particular hospital referent group is more likely to share common (i.e., mimetic) practices, leading each hospital to react similarly to peers within the group when responding to external pressures. However, having a high level of hospital geographical density within the referent group also may lead to more aggressive tactics to capture patient attention, than in a region with a low level of geographical density. For example, a hospital within a highly dense referent group area may face keen competition for the same types of patients. In this situation of high density, some hospitals within a referent group may be more likely to respond in appearance to external pressures (Marquis, 2003). Hospitals that are not capable of following the actual practices of the same geographic group are likely to adopt symbolic practice when responding to external pressures. In short, when hospitals have high referent group density, previously penalized hospitals are more likely to adopt symbolic practice. We posit the following hypothesis:

Hypothesis 2 (H2). *Compared to low-density hospital referent groups, symbolic management practice is more likely when penalized hospitals reside within a high-density referent group.*

4.2.3.3 Impact of Legitimacy Efforts on Symbolic Behavior

New hospital IT investments conforming to government mandates reflect an attempt to achieve institutionalization rather than symbolic practice, which we refer to using the term legitimacy efforts (DiMaggio and Powell, 1983). Prior studies in information systems (IS) research explore IT adoption and corresponding impacts on hospital operational performance (Agarwal et al., 2010). Modern hospitals incorporate IT to handle service processes smoothly (Angst and Agarwal, 2009). IT usage can enable process improvements and thus improve objective hospital performance, such as financial performance and quality performance (Devaraj et al., 2013), and enhanced process transparency (Kohli and Kettinger, 2004). Thus, we expect moderating effects of legitimacy efforts pertaining to IT adoption on hospital symbolic management practices. Even for hospitals previously having low VBP performance, in achieving the successful adoption of new government mandated IT systems, we expect those hospitals are more likely to improve actual process performance. In turn, these hospitals are less likely to exhibit characteristics consistent with symbolic practice. We posit the following hypothesis:

Hypothesis 3 (H3). *Legitimacy effort negatively moderates a previously penalized hospital's symbolic management practices.*

We also conjecture that IT adoption can moderate referent group social proximity effects upon symbolic management. Previous studies show diffusion effects of IT on hospitals (Angst et al., 2010), thus social proximity is positively associated with the likelihood of IT adoption, a form of mimetic isomorphism. When hospitals adopt new IT and implement it appropriately, such hospitals can deliver more effective healthcare services (Bardhan and Thouin, 2013). Thus, such outcomes are consistent with institutionalization rather than symbolic practices. We expect hospitals that adopt new IT in high-density areas are less likely to exhibit characteristic indicators of symbolic practice when responding to external government requirements. We posit the following hypothesis:

Hypothesis 4 (H4). *For hospitals within a dense referent group, when a hospital* *exhibits legitimacy effort, the impact diminishes on symbolic management practices.*

In summary, emerging healthcare literature on impacts of VBP pays little attention to the symbolic management perspective. From organizational literature, we observe many institutional pressures and corresponding responses. These studies seldom focus on strategic service design in the healthcare industry. We view the impact of institutional pressures on the healthcare industry as imperative research questions that have not been explored yet. Thus, we examine the actual effects of the government-initiated program (VBP) for both healthcare policy makers and hospital administrators.

4.3 Research Methodology

We next describe available data sources. We then describe how we construct variables pertaining to our analyses and econometric models.

4.3.1 Data Sources

The data for this study include hospital-level information related to hospital VBP performance and hospital operations. In particular, the data come from various sources, such as *Medicare Hospital Compare*, *CMS Cost Report*, *CMS Impact Files*, *CMS Case-Mix Index*, *Dartmouth Atlas*, and *Healthcare Information and Management Systems Society* (HIMSS). From the *CMS Impact Files* data, we obtain information regarding the hospital level annual performance of the VBP program from 2013 to 2014. There are 2,984 VBP participating acute care hospitals in 2013 and 2,728 hospitals in 2014. We use the *CMS Case-Mix Index* data to provide a signal about whether hospitals may not admit patients requiring complicated procedures. From *CMS Cost Report*, we obtain DSH information related to the proportion of patients eligible for extra financial bonuses to be paid to hospitals. From the *CMS Impact Files*, we obtain Outlier Payments information relevant to the percentage of patients that require extra costs to treat. Data from *Dartmouth Atlas* provides information pertaining to U.S. regional categorization (i.e., Hospital Referral Region (HRR)). This data set provides the number of care providers in a region, the level of hospital density, and other data, within each HRR. Using data from *HIMSS* Analytics, we observe the level of mandate-compliant IT adoption in each hospital. Specifically, this data set contains information on the adoption of Computerized Physician Order Entry (CPOE) technology. We collect information from *Medicare Hospital Compare* to use as treatment control variables. Specifically, we obtain patient satisfaction related data from HCHAPS, a part of the *Medicare Hospital Compare* data set. We also collect hospital-level control variables from *Medicare Hospital Compare*, including size of hospital (i.e., number of beds), teaching intensity, and hospital types. Table 4.2 illustrates definitions for our variables.

4.3.1.1 Dependent Variables

We use lagged dependent variables to assess the following year's hospital behaviors pertaining to symbolic practices. In response to the financial penalty, we estimate three dependent variables. From annual CMS Case-Mix Index (CMI) data, we observe the first symbolic behavior signal, whether a hospital decreases the CMI in the following year. Specifically, CMI is measured by summing the weighted treatment cases related to inpatient services and dividing by the number of cases (CMS, 2015a). Thus, a decreasing hospital case-mix indicates that a hospital treats less complicated patients compared to the previous year. To estimate the symbolic signal in each hospital i, we use the following formula:

$$\%\Delta CMI_i = \frac{CMI_{it+1} - CMI_{it}}{CMI_{it}}$$

Our second dependent variable is the change in the Disproportionate Share Hos-

Table 4.2: Variable definitions

	Variable Name	Variable Measure	Source
Dependent Variables	CMI	Percentage change of Case Mix Index from 2013 to 2014	*CMS Acute Inpatient PPS: Case-Mix Index*
	DSH	Difference between DSH percentage in 2013 and 2014	*CMS Cost Report*
	Outlier	Difference between operating and capital outlier payments as a percentage of the provider's Federal operating PPS payments in 2013 and 2014	*CMS Acute Inpatient PPS: Impact File*
Key Independent Variables	Penalty	Indicator variable: 1= hospital is penalized in 2013; 0 = otherwise	*CMS Hospital Value-Based Purchasing*
	Density	Indicator variable: 1= hospital is in the high-density area classified by HRRs; 0 = otherwise	*Dartmouth Atlas*
	CPOE	Indicator variable: 1= hospital mandated physicians to utilize CPOE; 0 = otherwise	*HIMSS Analytics*
Control Variables	ReadminFactor	Payment adjustment factor for the CMS readmissions penalty program	*CMS Acute Inpatient PPS: Impact File*
	Beds	Number of beds	*CMS Impact File*
	Revenue	Revenue from inpatient service (in $ 1,000,000)	*CMS Cost Report*
	Resident Bed Ratio	Resident to bed ratio in a hospital	*CMS Acute Inpatient PPS: Impact File*
	Region	Ten hospital regions defined by CMS	*CMS Acute Inpatient PPS: Impact File*
	Ownership	Ten hospital ownership types defined by CMS	*Medicare Hospital Compare: Hospital General Information*
Selection Models	Clean	VBP score of cleanliness and quietness (out of 10)	*Medicare Hospital Compare: HVBP-HCHAPS*
	Communication Doctor	VBP score of communication with doctors (out of 10)	*Medicare Hospital Compare: HVBP-HCHAPS*
	DischargeInfo	VBP score of providing discharge information (out of 10)	*Medicare Hospital Compare: HVBP-HCHAPS*
	Recommend	VBP score of patient recommended the hospital (out of 100)	*Medicare Hospital Compare: HVBP-HCHAPS*

pital (DSH) patient percentage, which measures the proportion of low-income and older patients (i.e., Medicare with Supplemental Security Income (SSI) patients or Medicaid patients) treated by the hospitals. Depending upon the DSH percentage, hospitals can achieve extra financial incentives, which can counteract the impact of the VBP penalty. A proportional change in DSH can further indicate another signal of symbolic practice by estimating whether a hospital tries to earn an extra financial revenue to compensate for a previous penalty. We use the percentage change in DSH as the second dependent variable.

$$\Delta\%DSH_i = \%DSH_{it+1} - \%DSH_{it}$$

The third dependent variable that we examine also pertains to hospital financial incentives. The Outlier Payments proportion indicates the percentage of patients who need unusually expensive treatments. Similar to DSH, if a hospital's percentage of Outlier Payments increases between years, the hospital can earn an additional financial revenue. Again, this revenue can compensate for a penalty. Thus, we use the percentage change in Outlier Payments as another signal of possible symbolic behavior by a hospital.

$$\Delta\%Outlier_i = \%OutlierPayment_{it+1} - \%OutlierPayment_{it}$$

4.3.1.2 Independent Variables

Independent variables pertaining to our research hypotheses reflect impacts of financial penalty, regional density, and the extent of IT adoption. The variable *Penalty* provides an indicator about whether a hospital was penalized due to prior VBP performance. *Penalty* is equal to 1 if a hospital was penalized and 0 otherwise. For region, we construct indicator variables for the hospital referral regions (HRRs), which are classified by the Dartmouth Atlas of Health Care, to account for the diffusion effects of the VBP program. HRRs delineate a regional hospital classification system for tertiary care (*www.DartmouthAtlas.org*). Using this information, we can capture the number of hospitals within the same HRR region as a specific hospital. We develop a binary variable *Density*, which is equal to 1 if a hospital is in an high-density area and 0 otherwise. We define high-density area based on the number of hospitals in a region being above the median number of hospitals across the HRRs. Thus, we capture the diffusion effects with a binary interaction term, *Penalty*Density*, which is equal to 1 for a penalized hospital in a dense HRR region and 0 otherwise.

To account for the moderating effects of government mandated information technology, we use the variable *CPOE* as an indicator to describe whether a hospital requires that physicians within a hospital utilize Computerized Physicians Order Entry (CPOE), which is the key measure for the Meaningful Use stage 1 (HealthIT, 2015). If a hospital adopts an IT system compliant with government mandates pertaining to information technology adoption (i.e., Meaningful Use), we capture the hospital as 1 and 0 otherwise.

4.3.1.3 Control Variables

We also measure several control variables that may influence our dependent variables. We control for hospital-level factors that include *Hospital Readmission Factor*, *Bed Size*, *Revenue*, *Cost*, and *ResidentBedRatio*. We also add regional demographic factors and hospital ownership.

4.3.1.4 Treatment Control Variables

The effect of the binary *Penalty* variable on the hospital behaviors can cause sample selection concerns, which occur due to different characteristics between the treated group (i.e., penalized) and the non-treated group. Thus, we consider several instrumental variables to account for the potential sample selection issues. *Clean* measures the patient satisfaction level for the hospital cleanliness and quietness. We expect a highly scored (out of 10) hospital is less likely to get penalized. *CommunicationDoctor* examines the quality of communication between patients and physicians in a hospital. If a hospital has a high quality of communication between patients and physicians, the hospital is less likely to be penalized. *DischargeInfo* represents whether a hospital provides appropriate discharge information to patients. If a hospital provides patients' discharge information clearly, the hospital may treat their patients sincerely, making the hospital less likely to be penalized. Finally, *Recommend* measures word of mouth in terms of patients who would recommend the hospital to their friends and relatives. Thus, the more the hospital is recommended by patients, the less the hospital is likely to be penalized.

Tables 4.3 and 4.4 show descriptive statistics and a correlation matrix for our data. We also checked for potential multicollinearity in the data using variance inflation factor (VIF). Since each VIF score does not exceed 5, and the average VIF is 2.51, we do not find evidence suggestive of multicollinearity.

4.3.1.5 Sample Sizes for Models

Out of 2,984 VBP participating hospitals, there are 1,774 hospitals for which we can obtain the percentage change in CMI information. After combining different data sets from various sources, we obtain 571 hospital observations that include all key independent variables for 2013 and 2014, such as Penalty, Density, and CPOE.

Table 4.3: Summary statistics

Variable	Mean	Std. Dev.	Min	Max
%ΔCMI	0.016	0.071	-0.376	2.015
Δ%DSH	0.006	0.035	-0.448	0.230
Δ%Outlier	0.003	0.107	-1.739	3.437
Penalty	0.478	0.5	0	1
Density	0.535	0.499	0	1
CPOE	0.55	0.498	0	1
ReadminFactor	0.997	0.004	0.9	1
Beds	186.30	179.78	1	1928
Revenue	243.59	514.02	0.0002	8440
Resident Bed Ratio	0.058	0.152	0	1.20
Clean	2.374	2.397	0	10
Communication Doctor	2.45	2.484	0	10
DischargeInfo	4.275	3.041	0	10
Recommend	61.2	26.117	0	100

Similarly, there are 802 hospitals for which we can obtain the change in DSH percentage information. Among them, only 275 hospitals provide full information relevant to our key measures, including Penalty, Density, and CPOE. Finally, 1,769 hospitals out of 2,984 VBP participating hospitals provide all relevant information pertaining to the change in outlier payment percentage. Among the 1,769 hospitals, 570 hospitals possess full information for Penalty, Density, and CPOE. Thus, the sample sizes of each model varies depending on the dependent variables, such that we have data on 571 hospitals for CMI, 275 hospitals for DSH, and 570 hospitals for Outlier payments.

4.3.2 Econometric Models

To examine whether symbolic practice may occur in hospitals in response to the VBP program, we consider how a penalty affects hospital behaviors. The treatment effect model can account for this causal effect. To describe our empirical strategy, let Y_i denote the dependent variable, such as hospital i's annual CMI percentage change in hospital i (%ΔCMI), change in DSH percentage (Δ%DSH), or the change in outlier payment percentage (Δ%Outlier). The general regression model is the

Table 4.4: Correlation matrix

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.
1.CMI	1.00													
2.DSH	0.12*	1.00												
3.Outlier	0.19*	0.15*	1.00											
4.Penalty	-0.09*	0.14*	0.02	1.00										
5.Density	-0.02	0.12*	-0.02	0.07*	1.00									
6.CPOE	0.08*	0.07*	0.07*	-0.07*	0.05	1.00								
7.ReadminFactor	0.18*	-0.07*	0.06*	-0.08*	-0.03	-0.04	1.00							
8.Beds	0.31*	0.26*	0.10*	0.11*	0.04*	0.09*	-0.14*	1.00						
9.Revenue	0.50*	0.21*	0.18*	0.04*	0.03	0.03	-0.07*	0.77*	1.00					
10.Resident Bed Ratio	0.28*	0.37*	0.15*	0.10*	0.10*	0.15*	-0.05	0.39*	0.35*	1.00				
11.Clean	-0.15*	-0.15*	-0.05*	-0.19*	-0.11*	-0.02	0.04	-0.33*	-0.33*	-0.20*	1.00			
12.Communication Doctor	-0.20*	-0.10*	-0.09*	-0.14*	-0.09*	-0.05	0.00	-0.32*	-0.37*	-0.15*	0.52*	1.00		
13.DischargeInfo	0.04*	-0.23	-0.02	-0.18*	-0.10	0.06	0.11*	-0.12*	-0.12	-0.03	0.22*	0.27*	1.00	
14.Recommend	0.45*	-0.24*	0.15*	-0.27*	-0.07*	0.12*	0.19*	-0.03	0.10*	0.00	0.31*	0.23*	0.37*	1.00

Notes. * indicates that coefficients are significant at the $p < 0.05$ level following:

$$
\begin{aligned}
Y_i = {} & \beta_0 + \beta_1 Penalty_i + \beta_2 Density_i + \beta_3 Penalty * Density_i \\
& + \beta_4 CPOE_i + \beta_5 Penalty * CPOE_i + \beta_6 Density * CPOE_i \\
& + \eta_1 ReadminFactor_i + \eta_2 log(Bed_i) + \eta_3 log(Revenue_i) + \eta_5 ResidentBedRatio_i \\
& + \gamma Region_i + \zeta Ownership_i + \epsilon_i,
\end{aligned}
\tag{4.1}
$$

where *Penalty* is the indicator of the treatment group (i.e., a hospital is previously penalized by the VBP program). Thus, this variable is the key estimator to measure H1. $Penalty * Density$ represents the diffusion effect of penalty used to assess H2. To assess H3 and H4, $Penalty * CPOE$ and $Density * CPOE$ are key variables to measure the impact of IT adoption in the hospital. $ReadminFactor$, *Bed*, *Revenue*, and *Cost* are control variables that may directly affect CMI change. Finally, we use ten *Region* dummy variables and ten *Ownership* dummy variables to control for hospital environments.

Following extant literature (Angrist, 2001; Greene, 2008), we develop instrumental variables to account for sample selection. Due to partial sample selection, $Penalty_i$ is likely to be correlated with ϵ_i. Since we cannot observe the difference in the expected value of penalty, versus the expected value of avoiding penalty for hospital i, we define another variable $Penalty^*$ as a latent variable.

$$
\begin{aligned}
& Penalty_i^* = \alpha_0 + \alpha_1 Clean_i + \alpha_2 CommunicationDoctor_i \\
& + \alpha_3 DischargeInfo_i + \alpha_4 Recommend_i + \nu_i \qquad (4.2) \\
& Penalty_i = 1 \ if \ Penalty_i^* > 0, \ Penalty_i = 0 \ otherwise
\end{aligned}
$$

where we assume error terms ϵ_i and ν_i have bivariate normal distributions with mean

0 and E(ϵ_i,ν_i) $\neq$ 0 for i. Using this instrumental variable approach, we account for selection issues and can perform consistent estimation. We use robust standard error estimates for the purpose of efficient estimation. In doing so, one can lessen worries about whether parameter estimates may be affected if our data set is not necessarily identically distributed (i.e., heteroskedasticity). We use SAS to prepare the data and estimate models using Stata version 12.

4.4 Empirical Findings

For each dependent variable, we estimate four models. Model 1 presents base models including estimates of main effects only. Model 2 focuses on H2 by including *Penalty*Density*. Model 3 is relevant to H3 by including *Penalty*CPOE*. Finally, Model 4 includes all relevant variables.

Since the sample selection model assumes a non-zero correlation (ρ) between a regression equation and a selection equation, we first checked ρ for our models. Using the Wald test of independence and inverse Mill's ratio (λ), we observed that all three main models (Model 4 in each table) have non-zero and significant ρ ($p < 0.05$). Thus, our sample selection assumption is appropriate. We also check the Wald χ^2 test of the regression model to estimate the goodness of fit and all models were statistically significant ($p < 0.01$). As a robustness check, we later present a propensity score matching analysis to demonstrate consistency of our results.

4.4.1 Econometric Results

Table 4.5 provides estimation results for the impact of financial penalty on hospital CMI behavioral changes. The coefficient of *Penalty* ($\beta = -0.039$, $p < 0.001$) is negative and significant, providing support for Hypothesis 1. We also observe a marginally significant interaction effect of CPOE IT adoption ($\beta = 0.012$, $p < 0.1$) indicating that symbolic practice is less when a hospital adopts government mandated IT (Hypothesis 3). However, we do not observe a signifiant moderating effect of penalty and density (β =−0.0011*, $p > 0.1$*, H2) or density and IT adoption (β =0.003*, $p > 0.1$*, H4) on CMI change.

Table 4.6 provides estimation results for the impact of financial penalty on hospital DSH changes. The coefficient of *Penalty* (β =0.017*, $p < 0.05$*) is positive and significant, providing support for Hypothesis 1. We also observe empirical evidence for the IT adoption variable (i.e., CPOE) indicating that adopting new IT is associated with increasing DSH percentage (β =0.009*, $p < 0.05$*). However, although we observe empirical evidence of the impact of financial penalty, we do not observe any signifiant moderating effects of density or IT adoption (H2 (β =0.004*, $p > 0.1$*), H3 (β =−0.007*, $p > 0.1$*), and H4 (β =0.004*, $p > 0.1$*)) with penalty on DSH change.

Table 4.7 provides estimation results for the impact of financial penalty on hospital Outlier payment changes. The coefficient of *Penalty* (β =0.044*, $p < 0.05$*) is positive and significant, providing support for Hypothesis 1. However, we do not observe any signifiant moderating effects of density or IT adoption (H2 (β =0.0001*, $p > 0.1$*), H3 (β =0.0024*, $p > 0.1$*), and H4 (β =0.001*, $p > 0.1$*)) with penalty on Outlier Payment change.

4.4.2 Discussion of Econometric Results

The estimated treatment effect (i.e., Penalty) indicates that other things being equal, penalized hospitals have a mean CMI percentage change that is 3.9% less than non-penalized hospitals. In other words, the penalized hospitals are less likely to admit the complicated patients than non-penalized hospitals. The difference is statistically significant at a 0.001 level. Similarly, other things being held equal, the penalized hospitals have a mean change in DSH percentage that is 1.2% greater than non-penalized hospitals. Also, the penalized hospitals have a mean change

Table 4.5: Estimation of the %ΔCMI Model

Regression Model	**Model1** Base Model-Penalty Only	**Model2** Penalty and Density	**Model3** Penalty and IT	**Model4** Penalty, Density, and IT
Penalty	-0.018^{*}	-0.02^{**}	-0.036^{**}	-0.039^{***}
	(0.007)	(0.007)	(0.01)	(0.001)
Density		0.00002		0.0026
		(0.003)		(0.006)
Penalty*Density		0.002		-0.0011
		(0.004)		(0.008)
CPOE			$-0.008^{\dagger}$	-0.007
			(0.005)	(0.005)
Penalty*CPOE			$0.013^{\dagger}$	$0.012^{\dagger}$
			(0.007)	(0.007)
Density*CPOE				0.003
				(0.006)
ReadminFactor	$-0.470^{\dagger}$	-0.713^{**}	-0.19	-0.53^{**}
	(0.245)	(0.230)	(0.27)	(0.21)
log(Beds)	-0.003	-0.003	-0.0001	0.0003
	(0.003)	(0.003)	(0.004)	(0.0047)
log(Revenue)	-0.0004	-0.0004	-0.001	-0.0017
	(0.001)	(0.001)	(0.002)	(0.0028)
ResidentBedRatio	0.016	0.023	0.30^{***}	0.366^{***}
	(0.08)	(0.09)	(0.083)	(0.081)
Region	Included	Included	Included	Included
Ownership	Included	Included	Included	Included
Selection Model				
Clean	-0.05^{**}	-0.05^{**}	-0.04	-0.04
	(0.015)	(0.015)	(0.026)	(0.026)
CommunicationDoctor	-0.033^{*}	-0.036^{*}	-0.076^{**}	-0.079^{**}
	(0.014)	(0.014)	(0.029)	(0.025)
DischargeInfo	-0.04^{***}	-0.04^{***}	-0.051^{**}	-0.049^{**}
	(0.01)	(0.01)	(0.020)	(0.019)
Recommend	-0.019^{***}	-0.019^{***}	-0.01	-0.009
	(0.004)	(0.004)	(0.007)	(0.007)
N	1774	1742	576	571
$\mathcal{X}^2$(df)	53.17(18)	58.38(20)	43.97(20)	68.73(23)
ρ	0.226	0.260	0.326	0.349
λ	0.01	0.01	0.014	0.015

$^{\dagger}\, p < 0.1$ $^{*}\, p < 0.05$ $^{**}\, p < 0.01$ $^{***}\, p < 0.001$

Table 4.6: Estimation of the Δ%DSH Model

Regression Model	**Model1** Base Model-Penalty Only	**Model2** Penalty and Density	**Model3** Penalty and IT	**Model4** Penalty, Density, and IT
Penalty	0.015^{**}	0.015^{**}	0.016^{*}	0.017^{*}
	(0.004)	(0.004)	(0.006)	(0.006)
Density		-0.0002		-0.0046
		(0.003)		(0.0039)
Penalty*Density		0.001		0.004
		(0.004)		(0.005)
CPOE			0.007^{*}	0.009^{*}
			(0.003)	(0.004)
Penalty*CPOE			-0.006	-0.007
			(0.005)	(0.006)
Density*CPOE				0.004
				(0.005)
ReadminFactor	0.002	-0.05	-0.58	-0.13
	(0.216)	(0.30)	(0.58)	(0.16)
log(Beds)	-0.001	-0.002	-0.006	-0.005
	(0.002)	(0.002)	(0.004)	(0.004)
log(Revenue)	0.003	0.003	0.003	0.003
	(0.001)	(0.001)	(0.002)	(0.002)
ResidentBedRatio	0.083	0.09	-1.38	Omitted
	(0.30)	(0.317)	(1.40)	
Region	Included	Included	Included	Included
Ownership	Included	Included	Included	Included
Selection Model				
Clean	-0.09^{***}	-0.09^{***}	-0.09^{*}	-0.08^{**}
	(0.022)	(0.023)	(0.03)	(0.03)
CommunicationDoctor	-0.025	-0.032	-0.067^{*}	-0.069^{**}
	(0.02)	(0.02)	(0.034)	(0.034)
DischargeInfo	-0.033^{*}	-0.030^{*}	-0.03	-0.033
	(0.016)	(0.017)	(0.028)	(0.028)
Recommend	-0.026^{***}	-0.024^{***}	-0.017	-0.016
	(0.006)	(0.007)	(0.012)	(0.012)
N	802	784	279	275
$\mathcal{X}^2$(df)	127.63(18)	89.34(20)	90.93(20)	4424.04(22)
ρ	-0.27	-0.28	-0.28	-0.28
λ	-0.007	-0.007	-0.006	-0.006

$^{\dagger}p < 0.1$ $^{*}p < 0.05$ $^{**}p < 0.01$ $^{***}p < 0.001$

Table 4.7: Estimation of the Δ%Outlier Payment Model

Regression Model	**Model1** Base Model-Penalty Only	**Model2** Penalty and Density	**Model3** Penalty and IT	**Model4** Penalty, Density, and IT
Penalty	-0.003	-0.003	0.0019	0.044*
	(0.008)	(0.008)	(0.0058)	(0.019)
Density		-0.0019		-0.0027
		(0.0019)		(0.029)
Penalty*Density		0.001		0.0001
		(0.0029)		(0.0036)
CPOE			-0.0037	-0.0039
			(0.0027)	(0.003)
Penalty*CPOE			0.0019	0.0024
			(0.005)	(0.005)
Density*CPOE				0.001
				(0.004)
ReadminFactor	-0.017	-0.075	0.060	0.140
	(0.088)	(0.107)	(0.125)	(0.156)
log(Beds)	-0.0006	-0.0008	0.002	0.025
	(0.0027)	(0.0027)	(0.004)	(0.004)
log(Revenue)	-0.0003	. -0.0001	-0.0014	-0.0013
	(0.0014)	(0.0014)	(0.002)	(0.002)
ResidentBedRatio	0.070***	0.073***	0.132***	0.149***
	(0.013)	(0.013)	(0.038)	(0.037)
Region	Included	Included	Included	Included
Ownership	Included	Included	Included	Included
Selection Model				
Clean	-0.049**	-0.048**	-0.025	-0.023
	(0.015)	(0.015)	(0.02)	(0.02)
CommunicationDoctor	-0.033*	-0.037*	-0.04	-0.04
	(0.016)	(0.015)	(0.024)	(0.024)
DischargeInfo	-0.04***	-0.04***	-0.045*	0.045*
	(0.011)	(0.011)	(0.019)	(0.018)
Recommend	-0.02***	-0.019***	-0.004	-0.004
	(0.005)	(0.004)	(0.007)	(0.007)
N	1769	1737	575	570
$\mathcal{X}^2$(df)	61.62(18)	63.25(20)	68.86(20)	81.43(23)
ρ	0.077	0.079	-0.76	-0.76
λ	0.002	0.002	-0.031	-0.031

$^{\dagger}\, p < 0.1$ $^{*}\, p < 0.05$ $^{**}\, p < 0.01$ $^{***}\, p < 0.001$

in Outlier Payment percentage that is 4.4% greater than non-penalized hospitals. Both differences are statistically significant at a 0.05 level. Thus, the results indicate that penalized hospitals may try to earn extra financial incentives by admitting more patients who can bring in such financial benefits. Overall, the results exhibit empirical evidence that the financially penalized hospitals use tactics consistent with symbolic practices, which may be unintended outcomes from the VBP objective.

In terms of moderating effects, the findings do not support the hypothesized moderating effects of density among the referent group of hospitals. In other words, we observe no managerial difference of penalized hospital behaviors pertaining to symbolic practices between an high density area and a low density area. Regarding the adoption of new government mandated IT systems compliant with Meaningful Use (i.e., CPOE), new hospital IT investments conforming to Meaningful Use government mandates in some cases appear to mitigate the characteristics consistent with symbolic practice. Specifically, in hospitals that are adopting the new IT systems, the mean CMI percentage change might lessen by 1.2% on average.

Overall, the estimation provides initial empirical evidence regarding our main research question: financially penalized hospitals appear more likely to exhibit characteristics consistent with adopting symbolic management practices. However, we observe no empirical evidence related to the moderating effects of density. Finally, we also observe that the moderating effects of IT adoptions are mixed depending on the dependent variables.

4.4.3 Robustness Considerations

We adopt propensity score matching analysis to ensure consistency of our empirical findings. A propensity score is a balanced one-dimensional score, which represents a vector of covariates (Rosenbaum and Rubin, 1983). Thus, this method enables one to examine an unbiased average treatment effect of the treated group (i.e., financially penalized hospitals) against the control group (i.e., not penalized hospitals). To estimate the impact of the penalty, we first need to obtain the propensity score. Following the propensity analysis literature (Guo and Fraser, 2010), we obtain the propensity score (predicted logit) via logistic regression. We used nearest neighbor matching to identify reasonable pairs of hospitals for comparison. Finally, each pair within the sample, which includes one treated hospital and one control hospital, provides the basis for examining the average treatment effect.

Table 4.8 provides estimation results of average treatment effect via propensity score matching. The significant average treatment effects for CMI indicate that penalized hospitals decrease by more than 1.4% CMI change compare to those for non-penalized hospitals. This finding again suggests that penalized hospitals have a higher incidence of turning away patients with complex care needs. Similarly, the average treatment effect for DSH shows an impact of 1.35% increase of DSH, indicating again that a penalized hospital tends to change operations in a manner that increasingly serves patients who can bring in extra financial benefits, compared to unpenalized hospitals. These two treatment effects are statistically significant ($p<0.05$). However, the average treatment effect of Outlier Payment is not statistically significant. Overall, the average treatment effects in this propensity score matching analysis have the same direction as our main treatment effect model findings. Thus, the propensity score results further support the consistency of our estimation.

4.5 Conclusion

This essay highlights hospital practice and process changes in response to the financial penalty incentives put in place by the Value Based Purchasing program. Specifically, the empirical estimation within this study provides evidence in sup-

Table 4.8: Treatment effect of penalty using propensity score matching

	Penalized Hospital	**Not Penalized Hospital**	**Difference between treated and control group**
CMI			
Before Matching	0.0078	0.026	-0.0148
After Matching	0.0080	0.023	-0.0145
DSH			
Before Matching	0.0082	0.0001	0.0083
After Matching	0.0099	-0.0036	0.0135
Outlier Payment			
Before Matching	-0.0017	-0.0096	-0.0007
After Matching	-0.0022	-0.0001	-0.0023

port of our hypotheses delivered from the related theoretical framework, symbolic management, which has yet to be quantitatively explored in much detail by OM researchers in the healthcare domain. In comparison to non-penalized hospitals, the financially penalized hospitals changed their process or practices in a manner that led to symptoms suggestive of symbolic management practices, such as declining patients who require more complicated care procedures that may lead to undesirable care outcomes, or accepting more patients who can provide an additional financial incentive. In general, our findings provide evidence that the VBP program may trigger penalized hospitals to conduct unintended behaviors, with respect to the underlying objective of the VBP program. In other words, hospitals are more likely to practice symbolically to avoid future financial penalty. Thus, the VBP program may result in unexpected outcomes.

Beyond the main effects, the study shows that the moderating effects vary depending upon the variables examined. The diffusion effect is not associated with any dependent variables. Our findings support the moderating effect of IT adoption only in the CMI change. The result is consistent with the extant IT literature that symbolic practice reduces when a hospital follows the government mandated IT system

(Devaraj et al., 2013). In summary, our empirical findings provide implications for rigorous research on hospitals' behaviors when they face external pressures.

4.5.1 Limitations

While this essay provides potentially meaningful implications for both academic scholars and practitioners, several limitations of the study are worth mentioning. First, the essay estimated impacts of the VBP policy across two time periods using cross sectional analysis. Although the findings provide empirical evidence to identify recent managerial tactics consistent with symbolic practices, the findings cannot address what will happen across multiple periods. Thus, research should continue to analyze the VBP policy. Specifically, scholars should seek to propose additional relevant research questions to examine other aspects of symbolic practices. For example, researchers might study whether all hospitals will eventually conform to VBP objectives, or whether an on-going exchange of outcomes happens across multiple periods between penalized and non-penalized hospitals. By accommodating multiple periods, researchers can address whether symbolic practices are transient phenomenon for hospitals, or a long-term issue. In other words, will government and other TPPs eventually figure out the practices to identify symbolic management, thus limiting it to being a short-run phenomenon?

Second, there may exist potential confounding issues related to each hospital's exogenous population change over time. For example, there might be overlap between CMI, DSH, and Outlier changes in a hospital. Thus, research should examine alternate dependent variables. One alternative way to construct a symbolic practice variable is to put together the three different outcomes (i.e., a variable is equal to 1 if a hospital decreases $\%\Delta$CMI, increases $\Delta\%$DSH, and increases $\Delta\%$Outlier), which may provide a more robust variable. In addition, future work should carefully control for other key variables. For instance, there will be different ways of controlling for the referent group. If future work considers such issues using other metrics, research may be able to identify significant moderating effects caused by characteristics of the referent groups. Possible applicable measures could include the average distance among hospitals, number of hospitals per mile, or whether a hospital is located in a big metro area. By doing so, we can provide more rigorous and consistent empirical findings.

Third, the study adopted the Heckman treatment effect model and the propensity score matching analysis to analyze the impact of a penalty on hospital behaviors. Although the proposed econometric approaches can reasonably address sample selection issues and potential inconsistent estimation results, the underlying phenomenon can also be addressed using other econometric approaches, such as the difference-in-difference model (DID), or hierarchical linear models (HLM). However, the current data set cannot be analyzed via the DID model due to the lack of multiple periods, and cannot employ HLM due to the lack of clear constructs that can define hierarchically nested sets. Thus, this essay can be extended when other testable data are available. The current essay also did not formally consider possible reverse causality. Future work should consider this issue to ensure consistent estimation.

Finally, the study used data from more than 500 hospitals for CMI analyses, 250 hospitals for DSH analyses, and 500 hospitals for Outlier analyses. A limitation pertains to the number of observations lost, relative to the 2984 VBP participating hospitals. The main reason for the loss of data is related to the available information about the CPOE variable from HIMSS Analytics. Thus, future studies should identify other IT variables that can avoid this loss of data, yet reasonably measure the same legitimacy effort phenomenon. Doing so should provide more consistent and reliable quantitative results to practitioners and academic scholars.

4.5.2 Discussion

The study contributes by developing a useful theoretical framework to delineate operational responses to the VBP program. This chapter empirically explores hospital and government incentive alignment and coordination problems in the healthcare supply chain context. By empirically quantifying characteristics consistent with symbolic practices of hospitals, the study transfers the institutional and symbolic management perspective into healthcare service operations management research. Next, there is little empirical research in healthcare operations management that examines the impact of VBP on hospitals' actual behaviors, thus this essay contributes quantitative findings pertaining to the VBP implementation. Since the objective of VBP is to achieve patient-oriented service delivery, which is an important topic in healthcare OM (Green, 2012), this research is in line with the same direction with the OM community. Thus, our study contributes by exploring the future research focus in operations management.

The findings from this study suggest several implications. First, healthcare OM scholars can extend the framework and findings of this study to examine symbolic management actions across the healthcare industry. For example, scholars interested in recent operating incentive programs, such as the bundled payment program (BPCI), could examine the process changes taking place in hospitals after managers choose to participate in the bundled payment program. As in related organizational theory literature (Elsbach et al., 1998; Fiss and Zajac, 2006; Rogers et al., 2007; Westphal and Graebner, 2010), researchers can also explore more rigorous OM related symbolic practice measures in healthcare organizations. Since there is no clear measure for examining symbolic practice, future research should make efforts to enhance measurement approaches and better identify the degree of symbolic practices in healthcare organizations.

Second, due to the lack of data availability, our study is based on an analysis of data from two time periods. Thus, future research can examine the VBP program using panel data with more than three time periods once such data becomes available. Future studies might benefit from adapting our theoretical framework on the impact of VBP financial penalties to other government incentives. In addition, future research might expand their data sets to include data at the department level or patient level. By doing so, researchers could gain generalizability and triangulation of findings.

To the best of our knowledge, this study is one of the few empirical OM research studies that relate to healthcare reimbursement processes. Thus, using our study as a first stepping-stone, researchers might provide substantial contributions by expanding their research interests into healthcare reimbursement processes, which has yet to be examined in detail.

For practitioners, our findings can provide important insights and implications for policy makers. Unlike other studies, our findings suggest implications relevant to potential responses to the VBP policy. In providing the empirical findings, government policy makers may review the current VBP program with a different lens.

5. CONCLUSION

This dissertation is inspired by managerial problem-motivated research, examining timely and critical issues for both academic researchers and healthcare practitioners. Many healthcare providers (e.g., hospital systems) generously share resources, such as operational information, through the academic-practice aligned healthcare research process. This type of research process helped motivate my individual dissertation chapters that comprise important but unexplored managerial questions. The dissertation chapters includes three essays involving (1) conceptual frameworks of the healthcare system and healthcare reimbursement processes, (2) analytical block scheduling models of outpatient scheduling processes, and (3) econometric models for hospital procurement behaviors in response to government financial penalties. Therefore, the dissertation covers various perspectives at the strategic level, process level, and organizational level.

The empirically grounded frameworks developed in Chapter 2 describe the complexity of healthcare reimbursement processes. The essay further explores how reimbursement processes create operating challenges, identifies research gaps, and provides suggestions for OM/SCM researchers to tackle these gaps. In Chapter 3, the Veterans Affairs scandal motivated me to consider improving outpatient scheduling. My work develops a scheduling algorithm for outpatient scheduling managers who face patient heterogeneity and patient no-shows, inspired by best practice scheduling approaches from the Toyota Production System. Chapter 4, inspired by government policies, concerns how managerial and operational responses of hospitals differ in response to external pressures imposed upon them by government financial incentive and quality improvement policies. The gap in both government expectations and hospital responses generated the impetus for an interesting empirical study into healthcare operations management.

Using multiple research methods, including grounded theory development for the first essay, deterministic analytical modeling for the second essay, and applied econometric analysis for the third essay, my dissertation provides both data and methodological triangulation. As described throughout the dissertation, modern healthcare operations management continues to increase in complexity. I hope my conceptual research frameworks, proposed scheduling modeling approaches, and corresponding theoretical framework and empirical findings, shed light on better healthcare process understanding pertaining to financial issues and impacts affecting healthcare OM decisions.

Yet admittedly, the findings described in this dissertation are not exhaustive. Thus, it is imperative to further discover related issues, such as healthcare reimbursement processes, the adoption of financial incentive programs, and more sophisticated scheduling algorithms. For example, in my future research I hope to discover operational impacts of process improvement initiatives related to financial flows that take place within a hospital at a department level, which has not been examined yet. Also, the adoption of information technologies in healthcare operations has a critical role in healthcare quality improvements. Thus, another research extension relates to the interface between information technology and financial flows in healthcare operations management. In line with my analytical scheduling essay, future research can be extended toward outpatient clinic overbooking policy to examine whether the block scheduling approach can lead to further improvements in outpatient scheduling.

REFERENCES

AAPS (2000). *AAPS report on Medicare fraud*. Association of American Physicians and Surgeons. (Accessed February 5, 2016) [http://www.aapsonline.org/fraud/medfraud.htm].

Abbey, D. (2009). *Healthcare Payment Systems An introduction*. CRC Press, Boca Raton, FL.

——— (2011). *Fee Schedule Payment System*. CRC Press, Boca Raton, FL.

——— (2012). *Prospective Payment Systems*. CRC Press, Boca Raton, FL.

——— (2013). *Cost-Based, Charge-Based, and Contractual Payment Systems*. CRC Press, Boca Raton, FL.

Adamy, J. (2014). *Report raises red flags on medicare lab billing*. The Wall Street Journal, July 9. (Accessed February 12, 2016) [http://www.wsj.com/articles/report-raises-red-flags-on-medicare lab-billing-1404878463].

Agarwal, R., Gao, G., DesRoches, C., Jha, A. (2010). *Research commentary-the digital transformation of healthcare: Current status and the road ahead*. Information Systems Research, 21(4):796–809.

AHA (2013). *Exploring the impact of the rac program on hospitals nationwide*. Results of AHA RAC TRAC Survey, 1st Quarter 2013. (Accessed February 12, 2016) [http://www.aha.org/content/13/12Q4ractracresults.pdf].

AMA (2013). *CPT 2013 Professional Edition*. American Medical Association, Atlanta, GA.

Amundson, S. (1998). *Relationships between theory-driven empirical research in operations management and other disciplines*. Journal of Operations Management, 16(4):341–359.

Angrist, J. (2001). *Estimation of limited dependent variable models with dummy endogenous regressors*. Journal of Business & Economic Statistics, 19(1):2–28.

Angst, C., Agarwal, R. (2009). *Adoption of electronic health records in the presence of privacy concerns: The elaboration likelihood model and individual persuasion*. MIS quarterly, 33(2):339–370.

Angst, C., Agarwal, R., Sambamurthy, V., Kelley, K. (2010). *Social contagion and information technology diffusion: the adoption of electronic medical records in US hospitals*. Management Science, 56(8):1219–1241.

Angst, C., Devaraj, S., Queenan, C., Greenwood, B. (2011). *Performance effects related to the sequence of integration of healthcare technologies*. Production and Operations Management, 20(3):319–333.

ASCA (2013). *What is an ASC?*. ASC Report. (Accessed February, 12) [http://www.ascassociation.org/AdvancingSurgicalCare/aboutascs/industryoverview].

Ata, B., Killaly, B., Olsen, T., Parker, R. (2013). *On hospice operations under Medicare reimbursement policies*. Management Science, 59(5):1027–1044.

Bach, P. (2009). *Limits on Medicare's ability to control rising spending on cancer drugs*. New England Journal of Medicine, 360(6):626–633.

Bailey, N. (1952). *A study of queues and appointment systems in hospital out-patient departments, with special reference to waiting-times*. Journal of the Royal Statistical Society, 14(2):185–199.

Baker, D., Rosnick, D. (2005). *The burden of social security taxes and the burden of excessive health care costs*. Center for Economic and Policy Research. (Accessed February 12, 2016) [http://cepr.net/publications/reports/the-burden-of-social-security

-taxes-and-the-burden-of-excessive-health-care-costs].

Bardhan, I., Oh, J., Zheng, Z., Kirksey, K. (2014). *Predictive analytics for readmission of patients with congestive heart failure*. Information Systems Research, 26(1):19–39.

Bardhan, I., Thouin, M. (2013). *Health information technology and its impact on the quality and cost of healthcare delivery*. Decision Support Systems, 55(2):438–449.

Barratt, M., Choi, T. (2007). *Mandated RFID and institutional responses: Cases of decentralized business units*. Production and Operations Management, 16(5):569–585.

Barton, P. (2010). *Understanding the US health services system*. Health Administration Press, Chicago, IL.

Benner, M., Tushman, M. (2003). *Exploitation, exploration, and process management: The productivity dilemma revisited*. Academy of Management Review, 28(2):238–256.

Berenson, R., Rich, E. (2010). *Us approaches to physician payment: The deconstruction of primary care*. Journal of General Internal Medicine, 25(6):613–618.

Bernard, T. (2012). *Discrepancies on medical bills can leave a credit stain*. The New York Times, May 4. (Accessed February 12, 2016) [http://www.nytimes.com/2012/05/05/your-money/medical-debts-can-leave-stains-on-credit-scores.html?=0].

Bhakoo, V., Choi, T. (2013). *The iron cage exposed: Institutional pressures and heterogeneity across the healthcare supply chain*. Journal of Operations Management, 6(31):432–499.

Bosch, P., Dietz, C. (2000). *Minimizing expected waiting in a medical appointment system*. IIE Transactions, 32(9):841–848.

Boudreau, J., Hopp, W., McClain, J., Thomas, L. (2003). *On the interface between* *operations and human resources management*. Manufacturing & Service Operations Management, 5(3):179–202.

Boyer, K., Gardner, J., Schweikhart, S. (2012). *Process quality improvement: an examination of general vs. outcome-specific climate and practices in hospitals*. Journal of Operations Management, 30(4):325–339.

Bozarth, C., Warsing, D., Flynn, B., Flynn, E. (2009). *The impact of supply chain complexity on manufacturing plant performance*. Journal of Operations Management, 27(1):78–93.

Brill, S. (2015). *What I learned from my $190,000 open-heart surgery*. Time, 185(1).

Buerhaus, P., Staiger, D., Auerbach, D. (2000). *Implications of an aging registered nurse workforce*. The Journal of the American Medical Aossociaton, 283(22):2948–2954.

Buntin, M., Burke, M., Hoaglin, M., Blumenthal, D. (2011). *The benefits of health information technology: A review of the recent literature shows predominantly positive results*. Health Affairs, 30(3):464–471.

Business Data Codes (2015). *NAICS 6214 Outpatient Care Centers*. Business Data Codes. (Accessed February 12, 2016) [`http://businessdatacodes.com/naics/code-6214-outpatient-care-centers`].

Cachon, G. (2003). *Supply chain coordination with contracts*. Handbooks in Operations Research and Management Science, 11:227–339.

Cayirli, T., Veral, E. (2003). *Outpatient scheduling in health care: A review of literature*. Production and Operations Management, 12(4):519–549.

Cayirli, T., Veral, E., Rosen, H. (2006). *Designing appointment scheduling systems for ambulatory care services*. Health Care Management Science, 9(1):47–58.

Cayirli, T., Yang, K., Quek, S. (2012). *A universal appointment rule in the presence of no-shows and walk-ins*. Production and Operations Management, 21(4):682–

697.

CDC, Centers for Disease Control, Prevention (2010). *National hospital ambulatory medical care survey: 2010 outpatient department summary tables*. (Accessed February 12, 2016) [http://www.cdc.gov/nchs/data/ahcd/nhamcs-emergency/2010-ed-web-tables.pdf].

——— (2013). *Resources for entertainment education content developers: Preventive health care*. (Accessed February 12, 2016) [http://www.cdc.gov/healthcommunication/toolstemplates/entertainmented/tips/preventivehealth.html].

Chandrasekaran, A., Senot, C., Boyer, K. (2012). *Process management impact on clinical and experiential quality: Managing tensions between safe and patient-centered healthcare*. Manufacturing & Service Operations Management, 14(4):548–566.

Chen, D., Preston, D., Xia, W. (2013). *Enhancing hospital supply chain performance: A relational view and empirical test*. Journal of Operations Management, 31(6):391–408.

Chen, R., Robinson, L. (2014). *Sequencing and scheduling appointments with potential call-in patients*. Production and Operations Management, 23(9):1522–1538.

Choi, T., Eboch, K. (1998). *The TQM paradox: relations among TQM practices, plant performance, and customer satisfaction*. Journal of Operations Management, 17(1):59–75.

Clark, J., Huckman, R. (2012). *Broadening focus: Spillovers, complementarities, and specialization in the hospital industry*. Management Science, 58(4):708–722.

CMS, Centers for Medicare Medicaid (2013). *Transaction and Code Sets Standards*. (Accessed February 12, 2016) [https://www.cms.gov/Regulations -and-Guidance/HIPAA-Administrative-Simplification/

TransactionCodeSetsStands/index.html].

——— (2014a). *EHR inventive programs*. (Accessed February 12, 2016) [http://www.cms.gov/ Regulations-and-Guidance/Legislations/EHRIncen -tiveProgrmas/index.html].

——— (2014b). *Hospital value-based purchasing program*. (Accessed February 12, 2016) [https://www.cms.gov/ Medicare/Quality-Initiatives -Patient-Asse -ssment-Instruments/hospital-value-based-purchasing/index.html].

——— (2015a). *Acute Inpatient PPS: Case Mix Index*. (Accessed February 12, 2016) [https: //www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Acute-Inpatient-Files-for-Download-Items/CMS022630.html].

——— (2015b). *Rural Health Clinic*. (Accessed February 12, 2016) [http://www.cms.gov/Outreach-and-Education/edicare-Learning-Network -MLN/MLNProducts/RuralHlthClinfctsht.pdf].

Cox, T., Birchall, J., Wong, H. (1985). *Optimising the queuing system for an ear, nose and throat outpatient clinic*. Journal of Applied Statistics, 12(2):113–126.

Das, S., Yaylacicegi, U., Menon, N. (2011). *The effect of information technology investments in healthcare: A longitudinal study of its lag, duration, and economic value*. Engineering Management, IEEE Transactions on, 58(1):124–140.

David, E. (2014). *Health care and the $20,000 Bruise.*. The Wall Street Journal, September 2. (Accessed September 5, 2014) [http://www.wsj.com/articles /eric-michael-david-health-care-and-the-20-000-bruise-1409700173].

Devaraj, S., Kohli, R. (2003). *Performance impacts of information technology: Is actual usage the missing link?*. Management Science, 49(3):273–289.

Devaraj, S., Krajewski, L., Wei, J. (2007). *Impact of ebusiness technologies on* *operational performance: The role of production information integration in the supply chain*. Journal of Operations Management, 25(6):1199–1216.

Devaraj, S., Ow, T., Kohli, R. (2013). *Examining the impact of information technology and patient flow on healthcare performance: A theory of swift and even flow (TSEF) perspective*. Journal of Operations Management, 31(4):181–192.

DiMaggio, P., Powell, W. (1983). *The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields*. American Sociological Review, 48(2):1–14.

Ding, D. (2014). *The effect of experience, ownership and focus on productive efficiency: A longitudinal study of us hospitals*. Journal of Operations Management, 32(1):1–14.

Dobrzykowski, D., D., Saboori, Hong, P., Kim, S. (2014). *A structured analysis of operations and supply chain management research in healthcare (1982–2011)*. International Journal of Production Economics, 147:514–530.

Dobson, G., Hasija, S., Pinker, E. (2011). *Reserving capacity for urgent patients in primary care*. Production and Operations Management, 20(3):456–473.

Dobson, G., Pinker, E., Van Horn, R. (2009). *Division of labor in medical office practices*. Manufacturing & Service Operations Management, 11(3):525–537.

Doshi, A., Dowell, G., Toffel, M. (2013). *How firms respond to mandatory information disclosure*. Strategic Management Journal, 34(10):1209–1231.

eCFR, Electronic Code of Federal Regulations (2015). *Conditions for Medicare payment part 424*. (Accessed February 12, 2016) [http://www.ecfr.gov/cgi-bin/text-idx?rgn=div6&*node* =42 : 3.0.1.1.11.1].

Edelson, J. (2013). *A beginner's guide to the MMSEA*. American Bar Association. (Accessed February 12, 2016) [http://www.americanbar.org/content/newsletter/groups/labor-law/err-ezine/2013/sum2013/mmsea.html].

Elsbach, K., Sutton, R., Principe, K. (1998). *Averting expected challenges through anticipatory impression management: A study of hospital billing*. Organization Science, 9(1):68–86.

Erdogan, S., Gose, A., Denton, B. (2015). *On-line appointment sequencing and scheduling*. IIE Transactions, 47(11):1267–1286.

Feldman, J., Liu, N., Topaloglu, H., Ziya, S. (2014). *Appointment scheduling under patient preference and no-show behavior*. Operations Research, 62(4):794–811.

Festinger, L. (1954). *A theory of social comparison processes*. Human Relations, 7(2):117–140.

Field, R. (2008). *Why is health care regulation so complex?*. Pharmacy and Therapeutics, 33(10):607.

Fisher, M. (2007). *Strengthening the empirical base of operations management*. Manufacturing & Service Operations Management, 9(4):368–382.

Fisher, M., Ramdas, K., Ulrich, K. (1999). *Component sharing in the management of product variety: A study of automotive braking systems*. Management Science, 45(3):297–315.

Fiss, P., Zajac, E. (2004). *The diffusion of ideas over contested terrain: The (non)adoption of a shareholder value orientation among german firms*. Administrative Science Quarterly, 49(4):501–534.

——— (2006). *The symbolic management of strategic change: Sensegiving via framing and decoupling*. Academy of Management Journal, 49(6):1173–1193.

Flynn, B., Huo, B., Zhao, X. (1994). *A framework for quality management research and an associated measurement instrument*. Journal of Operations Management, 10(4):339–366.

——— (2010). *The impact of supply chain integration on performance: A contingency and configuration approach*. Journal of Operations Management, 28(1):58–71.

Frandsen, B., Powell, M., Rebitzer, J. (2015). *Persistently inefficient? Organizational fragmentation, coordination failures, and common agency problems in the US health care system.*. Working Paper. (Accessed March 27, 2016) [http://economics.mit.edu/files/10730].

Fuloria, P., Zenios, S. (2001). *Outcomes-adjusted reimbursement in a health-care delivery system*. Management Science, 47(6):735–751.

Gao, G., McCullough, J., Agarwal, R., Jha, A. (2012). *A changing landscape of physician quality reporting: Analysis of patients online ratings of their physicians over a 5-year period*. Journal of Medical Internet Research, 14(1):e38.

Goes, J., Park, S. (1997). *Interorganizational links and innovation: The case of hospital services*. Academy of Management Journal, 40(3):673–696.

Goh, J., Gao, G., Agarwal, R. (2011). *Evolving work routines: Adaptive routinization of information technology in healthcare*. Information Systems Research, 22(3):565–585.

Goldstein, S., Iossifova, A. (2012). *Ten years after: Interference of hospital slack in process performance benefits of quality practices*. Journal of Operations Management, 30(1):44–54.

Goodrick, E., Salancik, G. (1996). *Organizational discretion in responding to institutional practices: Hospitals and cesarean births*. Administrative Science Quarterly, pages 1–28.

Green, L. (2012). *The vital role of operations analysis in improving healthcare delivery*. Manufacturing & Service Operations Management, 14(4):488–494.

Green, L., Kolesar, P. (2004). *Anniversary article: Improving emergency responsiveness with management science*. Management Science, 50(8):1001–1014.

Green, L., Savin, S. (2008). *Reducing delays for medical appointments: A queueing approach*. Operations Research, 6(56):400–412.

Green, L., Savin, S., Murray, M. (2007). *Providing timely access to care: What is the right patient panel size?*. The Joint Commission Journal on Quality and Patient Safety, 14(4):488–494.

Greene, W. (2008). *Econometric Analysis*. Prentice Hall, Upper Saddle River, NJ.

Guerriero, F., Guido, R. (2011). *Operational research in the management of the operating theatre: a survey*. Health Care Management Science, 14(1):89–114.

Guo, S., Fraser, M. (2010). *Propensity score analysis: Statistical methods and analysis*. Sage, Thousand Oaks, CA.

Gupta, D., Denton, B. (2008). *Appointment scheduling in health care: Challenges and opportunities*. IIE Transactions, 40(9):800–819.

Gupta, D., Wang, W. (2011). *Handbook of Healthcare System Scheduling– Patient appointments in ambulatory care: Challenges and opportunities*. Springer, Boston, MA.

Handfield, R., Melnyk, S. (1998). *The scientific theory-building process: A primer using the case of TQM*. Journal of Operations Management, 16(4):321–339.

Hans, E., Van Houdenhoven, M., Hulshof, P. (2012). *Handbook of healthcare system scheduling– A framework for healthcare planning and control*. Springer, Boston, MA.

Hare, W., Alimadad, A., Dodd, H., Ferguson, R., Rutherford, A. (2009). *A deterministic model of home and community care client counts in british columbia*. Health Care Management Science, 12(1):80–98.

Hawkins, M. (2012). *A survey of America's physicians: Practice patterns and perspectives*. The Physicians Foundation, January 26. (Accessed February 12, 2016) [http://www.physiciansfoundation.org/uploads/Physicians-Foundation-2012-Biennial-Survey.pdf].

HCCI, Health Care Cost Institute (2010). *Health care cost and utilization re-*

port. (Accessed February 12, 2016) [http://www.healthcostinstitute.org/files/HCCI-HCCUR2010.pdf].

He, B., Dexter, F., Macario, A., Zenios, S. (2012). *The timing of staffing decisions in hospital operating rooms: incorporating workload heterogeneity into the newsvendor problem*. Manufacturing & Service Operations Management, 14(1):99–114.

Health Law Resources (2012). *Third party payer*. (Accessed February 12, 2016) [https://www.healthlawyers.org/hlresources/Health-Law-Wiki/Third-Party-Payor.aspx].

HealthIT (2015). *Computerized Physician Order Entry (CPOE) for Medication Orders*. (Accessed February 17, 2016) [https://www.healthit.gov/providers-professionals/achieve-meaningful-use/core-measures/cpoe-meaningful-use].

Heim, G., Sinha, K. (2002). *Service process configurations in electronic retailing: A taxonomic analysis of electronic food retailers*. Production and Operations Management, 11(1):54–74.

Herriott, S. (1999). *Reducing delays and waiting times with open-office scheduling*. Family Practice Management.

Heskett, J. (1986). *Managing in the service economy*. Harvard Business School Press Boston, MA.

Heugens, P., Lander, M. (2009). *Structure! agency!(and other quarrels): A meta-analysis of institutional theories of organization*. Academy of Management Journal, 52(1):61–85.

HHS (2015). *The Affordable Care Act is working*. U.S. Department of Health and Human Services. (Accessed February 12, 2016) [http://www.hhs.gov/healthcare/facts/factsheets/2014/10/affordable-care-act-is-working.html].

Hill, A., Collier, D., Froehle, C., Goodale, J., Metters, R., Verma, R. (2003). *Research* *opportunities in service process design*. Quality Control and Applied Statistics, 48(2):225–228.

Hillman, A., Dalziel, T. (2003). *Boards of directors and firm performance: Integrating agency and resource dependence perspectives*. Academy of Management Review, 28(3):383–396.

Hixon, A., Chapman, R., Nuovo, J. (1999). *Failure to keep clinic appointments: Implications for residency education and productivity*. Family Medicine, 31:627–630.

Ho, C., Lau, H. (1992). *Minimizing Total Cost in Scheduling Outpatient Appointments*. Management Science, 38(12):1750–1764.

Huang, Y., Hanauer, D. (2014). *Patient no-show predictive model development using multiple data sources for an effective overbooking approach*. Applied Clinical Informatics, 5:836–860.

Huang, Y., Verduzco, S. (2015). *Appointment template redesign in a womens health clinic using clinical constraints to improve service quality and efficiency*. Applied Clinical Informatics, 5:271–287.

Hulshof, P., Kortbeek, R., Boucherie, E., Hans, E., Bakker, P. (2012). *Taxonomic classification of planning decisions in health care: A structured review of the state of the art in or/ms*. Health Systems, 1(2):679–686.

Hussey, P., Eibner, C., Ridgely, M., McGlynn, E. (2009). *Controlling US health care spendingseparating promising from unpromising approaches*. New England Journal of Medicine, 361(22):2109–2111.

Institute of Medicine, Committee on Quality of Health Care in America (2001). *Crossing the quality chasm: A new health system for the 21st century*. National Academy Press, Washington, D.C.

Jack, E., Powers, T. (2004). *Volume flexible strategies in health services: A research* *framework*. Production and Operations Management, 13(3):230–244.

Jencks, S., Dobson, A. (1987). *Refining case-mix adjustment*. New England Journal of Medicine, 317(11):679–686.

Jha, A., Orav, E., Epstein, A. (2010). *The effect of financial incentives on hospitals that serve poor patients*. Annals of Internal Medicine, 153(5):299–306.

Jiang, H., Pang, Z., Savin, S. (2012). *Performance-based contracts for outpatient medical services*. Manufacturing & Service Operations Management, 14(4):654–669.

Kallinikos, J., Tempini, N. (2014). *Patient data as medical facts: Social media practices as a foundation for medical knowledge creation*. Information Systems Research, 25(4):817–833.

KC, D. (2013). *Does multitasking improve performance? Evidence from the emergency department*. Manufacturing & Service Operations Management, 16(2):168–183.

KC, D., Terwiesch, C. (2011). *The effects of focus on performance: Evidence from California hospitals*. Management Science, 57(11):1897–1912.

——— (2012). *An econometric analysis of patient flows in the cardiac intensive care unit*. Manufacturing & Service Operations Management, 14(1):50–65.

Kellogg, D., Nie, W. (1995). *A framework for strategic service management*. Journal of Operations Management, 13(4):323–337.

Kensling, B., Nissenbaum, D. (2014). *VA goal to slash wait times was 'unrealistic,' aide said*. The Wall Street Journal, May 23. (Accessed February 12, 2016) [http://online.wsj.com/news/articles/SB10001424052702303749904579580473122138420].

Klassen, K., Rohleder, T. (1996). *Scheduling outpatient appointments in a dynamic environment*. Journal of Operations Management, 14(2):83–101.

Klassen, R., Leitch, K., Hora, M. (2010). *Paediatric orthopaedic clinic at the children's hospital of Western Ontario*. Ivey Publishing, (9B08D001).

Klepper, B. (2011). *Fixing America's health care reimbursement system*. Kaiser Health News, March 3. (Accessed February 12, 2016) [http://khn.org/news/030-311klepper/].

Kodjababian, J. (2003). *Improving patient access and continuity of care: A successful implementation of open access scheduling*. The 131st Annual Meeting of APHA. (Accessed March 27, 2016) [https://apha.confex.com/apha/131am/techprogram/paper-62912.htm].

Kohli, R., Kettinger, W. (2004). *Informating the clan: Controlling physicians' costs and outcomes*. MIS Quarterly, pages 363–394.

Kohn, L., Corrigan, J., M., Donaldson (1999). *To Err is human: Building a safer health system*. National Academy Press, Washington, DC.

Kostova, T., Roth, K. (2002). *Adoption of an organizational practice by subsidiaries of multinational corporations: Institutional and relational effects*. Academy of Management Journal, 45(1):215–233.

Koufteros, X., Cheng, T., Lai, K. (2007). *"Black-box" and "Gray-box" supplier integration in product development: Antecedents, consequences and the moderating role of firm size*. Journal of Operations Management, 25(4):847–870.

LaGanga, L., Lawrence, S. (2007). *Clinic overbooking to improve patient access and increase provider productivity*. Decision Sciences, 38(2):251–276.

——— (2012). *Appointment overbooking in health care clinics to improve patient service and clinic performance*. Production and Operations Management, 21(5):874–888.

Lahiri, A., Seidmann, A. (2012). *Information hang-overs in healthcare service systems*. Manufacturing & Service Operations Management, 14(4):634–653.

Landro, L. (2013). *The biggest mistakes doctors make*. The Wall Street Journal, November 17. (Accessed February 12, 2016) [http://www.wsj.com/articles/SB10001424052702304402104579151232421802264].

Law, A. (2013). *Simulation Modeling and Analysis*, volume 5. McGraw-Hill, New York, NY.

Lee, D., Zenios, S. (2012). *An evidence-based incentive system for Medicare's end-stage renal disease program*. Management Science, 58(6):1092–1105.

Lee, H. (2002). *Aligning supply chain strategies with product uncertainties*. California Management Review, 44(3):105–119.

Liu, N., Ziya, S. (2014). *Panel size and overbooking decisions for appointment-based services under patient no-shows*. Production and Operations Management, 23(12):2209–2223.

Lounsbury, M. (2001). *Institutional sources of practice variation: Staffing college and university recycling programs*. Administrative Science Quarterly, 46(1):29–56.

——— (2007). *A tale of two cities: Competing logics and practice variation in the professionalizing of mutual funds*. Academy of Management Journal, 50(2):289–307.

Majd, S. (2015). *Why your doctor always keeps you waiting?*. BusinessInsider, January 31. (Accessed February 12, 2016) [http://www.businessinsider.com/why-your-doctor-is-always-late-2015-1].

Majerowicz, A. (2009). *CPT Code Updates for 2009*. Journal of AHIMA. (Accessed February 12, 2016) [http://journal.ahima.org/2009/01/13/2009-cpt-coding-update].

Markóczy, L., S., Li, Peng, M., Ren, B. (2013). *Social network contingency, symbolic management, and boundary stretching*. Strategic Management Journal, 34(11):1367–1387.

Marks, M., Mathieu, J., Zaccaro, S. (2001). *A temporally based framework and taxonomy of team processes*. Academy of Management Review, 26(3):356–376.

Marley, K., Collier, D., G., Meyer (2004). *The role of clinical and process quality in achieving patient satisfaction in hospitals*. Decision Sciences, 35(3):349–369.

Marquis, C. (2003). *The pressure of the past: Network imprinting in intercorporate communities*. Administrative Science Quarterly, 48(4):655–689.

Marquis, C., Glynn, M., Davis, G. (2007). *Community isomorphism and corporate social action*. Academy of Management Review, 32(3):925–945.

Martinez-Moyano, I., McCaffrey, D., Oliva, R. (2013). *Drift and adjustment in organizational rule compliance: Explaining the "Regulatory Pendulum" in financial markets*. Organization Science, 25(2):321–338.

May, J., Spangler, W., Strum, D., Vargas, L. (2011). *The surgical scheduling problem: Current research and future opportunities*. Production and Operations Management, 20(3):392–405.

McKone-Sweet, K., Hamilton, P., Willis, S. (2005). *The ailing healthcare supply chain: A prescription for change*. Journal of Supply Chain Management, 41(1):4–17.

Medicare (2014). *Durable medical equipment (DME) coverage* . (Accessed February 12, 2016) [`http://www.medicare.gov/coverage/durable-medical-equipment-coverage.html`].

——— (2015). *Home health compare glossary terms and definitions*. (Accessed February 12, 2016) [`http://www.medicare.gov/homehealthcompare/Resources/Glossary.html`].

Medicare Claims Processing Manual (2008). *Chapter2–Admission and Registration Requirements:90.6–Definition of Encounter*. (Accessed February 12, 2016) [`http://www.cms.gov/Regulations-and-Guidance/Guidance/Manuals

/downloads/clm104c02.pdf].

Meyer, J., Rowan, B. (1977). *Institutionalized organizations: Formal structure as myth and ceremony*. American Journal of Sociology, 83(2):340–363.

Meyer, S., Collier, D. (2001). *An empirical test of the causal relationships in the baldrige health care pilot criteria*. Journal of Operations Management, 19(4):403–426.

Miller, A., Tucker, C. (2009). *Privacy protection and technology diffusion: The case of electronic medical records*. Management Science, 55(7):1077–1093.

Mondschein, S., Weintraub, G. (2003). *Appointment policies in service operations: A critical analysis of the economic framework*. Production and Operations Management, 12(2):266–285.

Moore, C., Wilson-Witherspoon, P., Probst, J. (2001). *Time and money: Effects of no-shows at a family practice residency clinic*. Family Medicine, 33(7):522–527.

Mukhopadhyay, T., Singh, P., Kim, S. (2011). *Learning curves of agents with diverse skills in information technology-enabled physician referral systems*. Information Systems Research, 22(3):586–605.

Murray, M., Tantau, C. (1999). *Redefining open access to primary care*. Managed Care Quarterly, 7:45–55.

——— (2000). *Same-day appointments: Exploding the access paradigm*. Family Practice Management, 7(8):45–45.

Muthuraman, K., Lawley, M. (2008). *A stochastic overbooking model for outpatient clinical scheduling with no-shows*. IIE Transactions, 40(9):820–837.

Nair, A., Nicolae, M., Narasimhan, R. (2013). *Examining the impact of clinical quality and clinical flexibility on cardiology unit performancedoes experiential quality act as a specialized complementary asset?*. Journal of Operations Management, 31(7):505–522.

Naveh, E., Katz-Navon, T., Stern, Z. (2005). *Treatment errors in healthcare: A safety climate approach*. Management Science, 51(6):948–960.

O'Hara, K. (2012). *CPT Code Updates for 2012*. Journal of AHIMA. (Accessed February 12, 2016) [http://journal.ahima.org/2012/01/27/cpt-code-updates-for-2012].

OIG, Office of Inspector General (2014). *Third party resources unit*. (Accessed February 12, 2016) [http://oig.hhsc. state.tx.us/chiefcounsel/tpr/aspx].

Oliver, C. (1991). *Strategic responses to institutional processes*. Academy of Management Review, 16(1):145–179.

O'Malley, K., Cook, K., Price, M., Wildes, K., Hurdle, J., Ashton, C. (2005). *Measuring diagnoses: Icd code accuracy*. Health services research, 40(5p2):1620–1639.

Ono, T. (1988). *Toyota Production System: Beyond Large-Scale Production*. Productivity Press, Portland, OR.

Peng, M. (2003). *Institutional transitions and strategic choices*. Academy of Management Review, 28(2):275–296.

Pesata, V., Pallija, G., Webb, A. (1999). *A descriptive study of missed appointments: Families' perceptions of barriers to care*. Journal of Pediatric Health Care, 13(4):178–182.

Pfeffer, J. (1981). *Management as symbolic action: The creation and maintenance of organizational paradigms.*. Research in Organizational Behavior, 3(1):1–52.

Pfeffer, J., Salancik, G. (2003). *The external control of organizations: A resource dependence perspective*. Stanford University Press, Palo Alto, CA.

Powell, A., Savin, S., Savva, N. (2012). *Physician workload and hospital reimbursement: Overworked physicians generate less revenue per patient*. Manfacturing & Service Operations Management, 14(4):512–528.

Prescott, P., Ryan, J., Soeken, K., Castorr, A., Thompson, K., Phillips, C. (1991).

The patient intensity for nursing index: A validity assessment. Research in Nursing & Health, 14(3):213–221.

Qu, X., Rardin, R., Williams, J., Willis, D. (2007). *Matching daily healthcare provider capacity to demand in advanced access scheduling systems*. European Journal of Operational Research, 183(2):812–826.

Quan, H., Li, B., S., Duncan, Parsons, G., Nilsson, C., Alibhai, A., Ghali, W. (2008). *Assessing validity of ICD-9-CM and ICD-10 administrative data in recording clinical conditions in a unique dually coded database*. Health Services Research, 43(4):1424–1441.

RAC (2014). *RAC program overview*. (Accessed February 12, 2016) [`http://www.recoveryauditcontractors.org/RAC-Program-Overview.html`].

Rau, J. (2012). *Medicare moves to tie doctor's pay to quality and cost of care*. Washington Post, April 14. (Accessed February 12, 2016) [`http://www.washingtonpost.com/national/health-science/medicare-moves-to-tie-doctors-pay-to-quality-and-cost-of-care`].

——— (2013). *Nearly 1,500 hospitals penalized under Medicare program rating qualtiy*. The Kaiser Health News, November 14. (Accessed February 12, 2016) [`http://www.kaiserhealthnews.org/stories/2013/november/14/value-based-purchasing-medicare.aspx`].

Reichers, A. (1985). *A review and reconceptualization of organizational commitment*. Academy of Management Review, 10(3):465–476.

Robinson, L., Chen, R. (2010). *A comparison of traditional and open-access policies for appointment scheduling*. Manufacturing & Service Operations Management, 12(2):330–346.

Rogers, K., Purdy, L., Safayeni, F., Duimering, P. (2007). *A supplier development program: rational process or institutional image construction?*. Journal of Operations Management, 25(2):556–572.

Rosenbaum, P., Rubin, D. (1983). *The central role of the propensity score in observational studies for causal effects*. Biometrika, 70(1):41–55.

Rosenthal, M. (2008). *Beyond pay for performanceemerging models of provider-payment reform*. New England Journal of Medicine, 359(12):1197–1200.

Rosenzweig, E., Roth, A., Dean, J. (2003). *The influence of an integration strategy on competitive capabilities and business performance: An exploratory study of consumer products manufacturers*. Journal of Operations Management, 21(4):437–456.

Roth, A., Menor, L. (2003). *Insights into service operations management: a research agenda*. Production and Operations management, 12(2):145–164.

Roth, A., Van Dierdonck, R. (1995). *Hospital resource planning: concepts, feasibility, and framework*. Production and Operations Management, 4(1):2–29.

Rouse, W., Serban, N. (2014). *Understanding and Managing the Complexity of Healthcare*. MIT Press, Cambridge, MA.

Ruef, M., Scott, W. (1998). *A multidimensional model of organizational legitimacy: Hospital survival in changing institutional environments*. Administrative Science Quarterly, pages 877–904.

Rust, C., Gallups, N., Clark, W., Jones, D., Wilcox, W. (1995). *Patient appointment failures in pediatric resident continuity clinics*. Archives of Pediatrics & Adolescent Medicine, 149(6):693–695.

Schmenner, R. (2004). *Service businesses and productivity*. Decision Sciences, 35(3):333–347.

Schütz, H., Kolisch, R. (2013). *Capacity allocation for demand of different customer-product-combinations with cancellations, no-shows, and overbooking when there is a sequential delivery of service*. Annals of Operations Research, 206(1):401–423.

Scott, W., Ruef, M., Mendel, P., Caronna, C. (2000). *Institutional change and organizations: Transformation of a healthcare field*. The University of Chicago Press, Chicago, IL.

Shortell, S, O'Brien, J, Carman, J, Foster, R, Hughes, E, Boerstler, H., O'Connor, E. (1995). *Assessing the impact of continuous quality improvement/total quality management: Concept versus implementation.*. Health Services Research, 30(2):377.

Silver-Greenberg, J. (2011). *How to fight a bogus bill*. The Wall Street Journal, February 19. (Accessed February 12, 2016)[`http://www.wsj.com/articles/SB10001424052748703312904576146371931841968`].

Silverman, E., Skinner, J. (2004). *Medicare upcoding and hospital ownership*. Journal of Health Economics, 23(2):369–389.

Sinha, K., Kohnke, E. (2009). *Health care supply chain design: Toward linking the development and delivery of care globally*. Decision Sciences, 40(2):197–212.

Soeken, K., Prescott, P. (1991). *Patient intensity for nursing index: The measurement model*. Research in Nursing & Health, 14(4):297–304.

Song, H., Tucker, A., Murrell, K. (2015). *The diseconomies of queue pooling: An empirical investigation of emergency department length of stay*. Management Science, 61(12):3032–3053.

Soriano, A. (1966). *Comparison of two scheduling systems*. Operations Research, 14(3):388–397.

Spear, S. (2005). *Fixing health care from the inside, today*. Harvard Business Review, 83(9):78.

Suchman, M. (1995). *Managing legitimacy: Strategic and institutional approaches*. Academy of Management Review, 20(3):571–610.

Swink, M., Narasimhan, R., Wang, C. (2007). *Managing beyond the factory walls: Effects of four types of strategic integration on manufacturing plant performance*.

Journal of Operations Management, 25(1):148–164.

Taylor, B. (1984). *Patient use of a mixed appointment system in an urban practice*. British Medical Journal, 289(6454):1277–1278.

Taylor, H., Morrison, I. (2011). *The incredible and wasteful complexity of the US healthcare system*. The Healthcare Blog, March 24. (Accessed February 12, 2016) [http://thehealthcareblog.com/blog/2011/03/24/the-incredible-and-wasteful-complexity-of-the-us-healthcare-system/].

Tessman, R. (2014). *New membership offers unlimited primary care*. KARE11, October 14. (Accessed February 12, 2016) [http://legacy.kare11.com/story/news/health/2014/10/14/health-membership-unlimited-clinic-visits/17269665/].

Tricare (2013). *Medicare claims process*. (Accessed February 12, 2016) [http://www.tricare.mil/Resources/Claims.aspx].

Tucker, A. (2004). *The impact of operational failures on hospital nurses and their patients*. Journal of Operations Management, 22(2):151–169.

——— (2007). *An empirical study of system improvement by frontline employees in hospital units*. Manufacturing & Service Operations Management, 9(4):492–505.

Tucker, A., Nembhard, I., Edmondson, A. (2007). *Implementing new practices: An empirical study of organizational learning in hospital intensive care units*. Management Science, 53(6):894–907.

Virji, A. (1990). *A study of patients attending without appointments in an urban general practice.*. British Medical Journal, 301(6742):22–26.

Von Drehle, D. (2014). *Medicine is about to get personal*. Time, December 22. (Accessed February 12, 2016) [http://time.com/3643841/medicine-gets-personal/].

Voss, C. (1992). *Applying service concepts in manufacturing*. International Journal of Operations & Production Management, 12(4):93–99.

Walter, S. (1973). *A comparison of appointment schedules in a hospital radiology department.*. British Journal of Preventive & Social Medicine, 27(3):160–167.

Wang, P. (1997). *Optimally scheduling n customer arrival times for a single-server system*. Computers & Operations Research, 24(8):703–716.

Weldon, D. (2014). *4 hot healthcare topics for 2015: Tech, M&A, health exchanges, reimbursement. Healthcare Finance*. Healthcare Finance News, December 30. (Accessed February 12, 2016) [`http://www.healthcarefinancenews.com/news/4-hot-healthcare-topics-2015-tech-ma-health-exchanges-reimbursement`].

Werner, R., Dudley, R. (2012). *Medicares new hospital value-based purchasing program is likely to have only a small impact on hospital payments*. Health Affairs, 31(9):1932–1940.

Westphal, J., Graebner, M. (2010). *A matter of appearances: How corporate leaders manage the impressions of financial analysts about the conduct of their boards*. Academy of Management Journal, 53(1):15–44.

Westphal, J., Gulati, R., Shortell, S. (1997). *Customization or conformity? An institutional and network perspective on the content and consequences of TQM adoption*. Administrative Science Quarterly, pages 366–394.

Westphal, J., Zajac, E. (1994). *Substance and symbolism in ceos long-term incentive plan*. Administrative Science Quarterly, 39(3):367–90.

——— (1998). *The symbolic management of stockholders: Corporate governance reforms and shareholder reactions*. Administrative Science Quarterly, pages 127–153.

——— (2001). *Decoupling policy from practice: The case of stock repurchase programs*. Administrative Science Quarterly, 46(2):202–228.

White, M., Pike, M. (1964). *Appointment systems in out-patients' clinics and the effect of patients' unpunctuality*. Medical Care, 12(4):133–145.

WHO, World Health Organization (2009). *Conceptual framework for the international classification for patient safety*. Final Technical Report. (Accessed February 12, 2016) [http://www.who.int/patientsafety/taxonomy/icps-full-report.pdf].

Williams, M. (2015). *Online appointment scheduling expected to increase in healthcare*. Business Solutions, January 20. (Accessed February 12, 2016) [http://www.bsminfo.com/doc/online-self-appointment-scheduling-expected-to-increase-in-healthcare-0001].

WorldBank (2014). *Health expenditure data total*. GDP Report. (Accessed February 12, 2016) [http://data.worldbank.org/indicator/SH.XPD.TOTL.ZS].

Yaraghi, N., Du, A., Sharman, R., Gopal, R., Ramesh, R. (2014). *Health information exchange as a multisided platform: Adoption, usage, and practice involvement in service co-production*. Information Systems Research, 26(1):1–18.

Zacharias, C., Pinedo, M. (2014). *Appointment scheduling with no-shows and overbooking*. Production and Operations Management, 23(5):788–801.

Zajac, E., Westphal, J. (1995). *Accounting for the explanations of CEO compensation: Substance and symbolism*. Administrative Science Quarterly, 40(2):283–308.

APPENDIX A

ACRONYMS AND COMMON TERMINOLOGY

Table A.1: Acronyms and common terminology

Acronyms	**Terminology**	**Definition**
AMA	American Medical Association	A professional group of physicians in the United States including both doctors of medicine (MD), doctors of osteopathic medicine (DO), and medical students.
APCs	Ambulatory Payment Classifications	The code classification program for United States hospital outpatient services. APC is an example of a prospective payment system (PPS). APCs are applicable only for hospitals, not for physicians.
ASCs	Ambulatory Surgical Centers	Ambulatory surgery centers are health care providers where surgical procedures do not require an overnight stay. Such surgery is less complicated than general hospital surgical procedures.
CAH	Critical Access Hospital	Critical Access Hospitals are rural area hospitals that receive cost-based reimbursement. To be designated a CAH, a rural care provider should satisfy specific criteria (i.e., the Conditions of Participation (CoP) 42CFR485).
CCO	Chief Compliance Officer	The chief compliance officer of a healthcare provider manages regulatory compliance issues within a care provider. Specifically, the CCO oversees reimbursement processes.
Claim Adjudication		Claim Adjudication refers to the determination of a TPPs payment after a care providers insurance benefits are applied to a medical claim.
CMS	Centers for Medicare and Medicaid Services	The Centers for Medicare & Medicaid Services is a federal agency that manages the Medicare program and Medicaid programs in cooperation with state governments.
CORF	Comprehensive Outpatient Rehabilitation Facility	A Comprehensive Outpatient Rehabilitation Facility is a care provider that offers rehabilitation of a patients medical issues via outpatient diagnostic, therapeutic, and restorative services.

Table A.1 continued

Acronyms	Terminology	Definition
CPT	Current Procedural Terminology	The Current Procedural Terminology code set is a medical code classification managed by the AMA. The CPT code set describes evaluation, management, and actual surgical procedures for providing standard service procedures for coordination among physicians, coders, and third party payers for reimbursement purposes.
DME	Durable Medical Equipment	Durable Medical Equipment suppliers provide medical products that a doctor prescribes for patients to use in the patients home.
DRGs	Diagnosis-Related Groups	Diagnosis-Related Groups represent a prospective payment classification system for inpatient services for the purposes of payment.
ED	Emergency Department	A care provider, which provides acute care services for patients who admit without a prior appointment. In general, patients walk in by themselves or are presented by ambulances. See also ER for Emergency Room.
EHR	Electronic Health Record	A software system for collection of electronic patient health information.
EMTALA	Emergency Medical Treatment and Active Labor Act	EMTALA is legislation that requires care providers to provide emergency health care treatment to anyone needing it regardless of citizenship, legal status, or ability to pay (EMTALA 2014).
FQHC	Federally Qualified Health Center	FQHC provides grants to care providers that are qualified under CMS and the Public Health Service Act (PHS).
GSP	Global Surgical Package	GSP is single payment for a package of services, generally surgical services.
HCCI	Health Care Cost Institute	HCCI is a non-profit organization that supports healthcare reimbursement processes by drawing on health care cost and utilization data for U.S. patients covered by third party payers.
HCPCS	Healthcare Common Procedure Coding System	HCPCS is a set of medical care procedure codes related to CPT codes. This code set is promulgated and maintained by CMS.
HHAs	Home Health Agencies	HHA are care providers that deliver short-term skilled nursing or rehabilitative services to homebound patients following a decline in function or an acute illness.
HHAPPS	Home Health PPS	Prospective payment systems for HHAs.
HIE	Health Information Exchange	HIE is a networked software resource that allows care providers and patients to access and share a patients medical information electronically.

Table A.1 continued

Acronyms	Terminology	Definition
HIPAA	Health Insurance Portability and Accountability Act of 1996	HIPAA is a federal law to make it easier for people to keep health insurance, protect the confidentiality and security of healthcare information, and help the healthcare industry control administrative costs.
HIPAA TCS	HIPAA Transaction and Code Standards	In HIPAA regulations, TCS provides standard transactions for Electronic Data Interchange (EDI) of health care data.
HOPPS	Hospital Outpatient Prospective Payment System	HOPPS is a reimbursement PPS for hospital outpatient services.
ICD	International Classification of Diseases	ICD is the standard system adopted by the United States and other countries to classify codes for health conditions and related information.
LTCH	Long-Term Care Hospital	LTCHs are certified as acute care hospitals, but LTCHs focus on patients who, on average, stay more than 25 days.
Meaningful Use		Program that provides financial incentives for the meaningful use of EHR technology.
Medicaid		Medicaid is a healthcare program for people with low income. See also CMS.
Medicare		Medicare is a federal health insurance program for elderly people. See also CMS.
MMSEA	Medicare, Medicaid and SCHIP Extension Act of 2007	MMSEA is legislation that, among other features, enables CMS to manage reimbursement processes properly by determining the role of the primary payer and secondary payer in the healthcare system.
MSP	Medicare Secondary Payer	MSP is a program where Medicare does not have primary payer role in the healthcare reimbursement system, but Medicare is involved in the reimbursement process as a secondary payer or possible tertiary payer.
MPFS	Medicare Physician Fee Schedule	A fee schedule payment system that is used by Medicare based on RBRVS.
POC	Plan Of Care	Written physician or authorized care provider orders for services and treatments based on the patients condition. The physician establishes the boundaries of care throughout the duration of the treatment.
PPS	Prospective Payment System	PPS is a payment system in which the reimbursement process is based on fee for service, with payments fixed in advance, generally for a year.
PWP	Patient Web Portal	A website that provides patient personal health information such as recent doctor visits, discharge summaries, diagnostic tests, and other information.

Table A.1 continued

Acronyms	Terminology	Definition
RAC	Recovery Audit Contractors	RACs audit care providers to identify and correct improper payments made under Medicare. RBRVS Resource-Based Relative Value Scale Through the use of relative values, RBRVS determines how much money care providers should be reimbursed under a fee schedule payment system.
RHC	Rural Health Clinic	The Rural Health Clinic Services Act of 1977 was enacted to provide primary care services in approximately 4,000 locations for Medicare patients in rural areas.
SNF	Skilled Nursing Facilities	SNFs are care providers certified for delivering specialized services, such as rehabilitation and various medical and nursing procedures.
Subrogation		One TPP takes over the obligations of payment from another TPP under an agreement based on a contract between the two TPPs.
TPP	Third Party Payer	TPPs reimburse care providers instead of patients. Examples of TPPs are commercial insurance companies, Medicare, and Medicaid.
VBP	Value-Based Purchasing Program	A CMS incentive program that offers reimbursement incentives or penalties based on a hospitals performance across a set of healthcare outcome metrics.
WHO	World Health Organization	A specialized agency of the United Nations, established in 1948 to avoid the international spread of diseases.

APPENDIX B

TAXONOMY OF HEALTHCARE INTENSITY

To broaden and refine operations strategy in healthcare, researchers need appropriate frameworks that encompass the overall U.S. healthcare system, including reimbursement processes. Due to the lack of such a framework, we must provide a taxonomy of the U.S. healthcare system. The taxonomy can be widely applicable to healthcare stakeholders. To develop the taxonomy of healthcare intensity, we adapt Marks et al. (2001)'s taxonomy development protocol, which includes academic literature reviews and incorporating advices from healthcare professionals. Our research framework comprises a hierarchical structure. The level of intensity contains two categories: (1) patient intensity level and (2) care provider intensity level. These three categories include several subcategories. We argue that operations challenges and medical errors of all types diagnostic and administrative are more likely to arise when the intensity levels of patient to healthcare provider are mismatched. Table 2.8 provides our taxonomy with specific definitions. We describe each intensity dimension below.

Patient Encounter Intensity Patient encounter intensity refers to the level of patient intensity associated with consuming a healthcare providers resource, such as workloads of physicians. We use three measures to classify the intensity level of patients: severity of a medical issue, complexity of procedures, and duration of illness. Patient classification systems have been widely studied in the nursing industry because resource allocation problem in nursing care are significantly associated with patient groups (Prescott et al., 1991). Typically, several dimensions are associated with a patient classification. One way to classify and develop patient groups is the use of attributions of patients medical issues. Healthcare researchers and government agencies have provided that the volume or service, and the required skill levels to treat patients, can separate patient intensity level (Soeken and Prescott, 1991). In particular, the patient intensity level can be determined by the degree of harm, which refers to the rigorousness, and period of any harm (WHO, 2009). Thus, depending on the severity of a medical issues (skill levels), complexity of procedure (skill levels), or duration of illness (the volume of service), we can label four patient intensity groups: Minimal-intensity level, Low-intensity level, Mid-intensity level, and High-intensity level.

Minimal intensity patient This patient group refers to patients, who can conduct their daily living without any issues, even if they do not present to the healthcare provider (i.e., clinics or hospitals). Based on WHO (2009)'s degree of harm classification, we can allocate patients who have little detected symptoms to this group. This group includes the lowest severity level, so patient dependency needs are not required. The process and decision making complexity level for treating a patient are the simplest as compared to other patient groups. This patient group also requires the least hours of services actually provided. One simple example for no intensity level of the patient is the Grade 1 finger sprain case, stretching or micro tearing of finger ligament tissues (NYU Langone Medical Center 2013). Another example would be a patient encounter for an executive checkup, which entails blood testing, routine urine examination, or other simple physical examinations. Such examinations allow the patient to perform any daily activities with minimal impairment or additional treatment. As this is preventive care, the severity level is lower than any other patient groups (CDC and Prevention, 2013).

Low intensity patient The second type of patient group is the low intensity level of patients. This low intensity group contains patients who have mild symptoms of short duration of illness, but need treatment from healthcare providers to stay healthy and lead active lives (WHO 2009). Outpatients or home care patients can be located in this group. First, outpatient refers to the patient, who is served the same day medical treatment (Cayirli and Veral, 2003). Thus, the total time spent for treating an outpatient has low intensity level, and the severity and complexity of outpatient diseases are relatively simpler than in other patient groups except for the minimal intensity level of patients. A patient with the Grade-2 or 3 finger sprain, which has partial or severe tearing of ligament tissue (NYU Langone Medical Center 2013), can be an appropriate example of low intensity level of patients. While the Grade 1 figure sprain explains a stable joint, the Grade 2, or 3 finger sprain represents instability of the joint. Thus, without assistance from a professional, the Grade 2, or 3 finger sprain patient must have difficulties to do ordinary activities. A home care patient group who is treated medical intervention by licensed professionals (Buerhaus et al., 2000), is another example of low intensity level of patients. Since the complexity of procedure or severity level is not high enough to treat patients in hospitals, we can put home care patients at this low level intensity group.

Moderate intensity patient The third type of patient group represents the medium intensity patient group. In this group, severity of illness, or complexity of procedures requires higher than the low intensity level. Thus, patients in this group need additional medical intervention. In general, this type of patient should be hospitalized, and the length of stay demands more than one day (Hulshof et al., 2012). Thus, medium intensity level of patient needs longer periods of treatment provided than lower or minimal intensity level of patient. At the medium intensity level patients eventually need inpatient services, healthcare providers should prepare appropriate facilities, such as intensive care units, nursing rooms, or operating rooms (Guerriero and Guido, 2011). For example, a patient who developed appendicitis can be categorized into the medium intensity level. This patient may have difficulty in performing ordinary activities without complete care for his condition or disease. Thus, we can put some part of inpatient at this medium intensity level.

High intensity patient Finally, high intensity level of the patient requires major surgical or medical intervention, which relates to a serious or potentially fatal illness. This type of illness comes from either acute illness and trauma, or active chronic diseases that associate with mortality. For example, a patient with myocardial infarction can be a high intensity patient. This type of patient should be served immediately, such as through emergency care services. For these patients, the severity of illness is highest, surgical procedures are more complex, and durations of illness are longer than in any other patient groups. With high intensity level, patients are exposed to shortening life expectancy, or major surgical procedures (WHO 2009). Thus, health care providers offer both emergency care services, and surgical care services to address this type of patients (Hans et al., 2012)

Care Provider Intensity For the second taxonomy dimension, we define the care provider as an individual or organization that offers healthcare services to patients (Abbey 2009). This taxonomy dimension is developed by the degree of care providers capability to treat patients, and the degree of the payment process complexity. According to government agencies, such as the Centers for Medicare and Medicaid Services (CMS) or the U.S. Department of Health & Human Services (HSS), there exist hundreds of different types of care providers (CMS 855 Enrollment Forms, and National Provider Identifier Standard). These agencies assign formal structures, such as legitimacies. These legitimacies classify into various health care providers. For example, an organization that provides inpatient services, such as Hospitals, should deliver designated services associated with a length of stay or period of care. Thus, care provider intensity level is determined by the healthcare providers service characteristics. Previous literature in the operations management field discusses a different type of healthcare service typologies (Dobrzykowski et al., 2014; Hans et al., 2012; Hulshof et al., 2012). In particular, Hulshof et al. (2012)s taxonomy suggests six types of health care services: Ambulatory care services, Emergency care services, Surgical care services, Inpatient care services, Homecare services, and Residential care services. Adapted by this taxonomy, we develop four dimensions of care provider intensity groups: Minimal intensity level, Low intensity level, Medium intensity level, and High intensity level of care providers.

Minimal intensity care provider The minimal intensity level of care provider refers to the organization that provides tangible healthcare items (i.e., wheelchairs, walkers, canes, or pillows). Thus, this type of provider can be recognized as a durable medical equipment (DME) supplier named by the Medicare program (Medicare, 2014). In this group, care providers sell healthcare items to patients and then are reimbursed under the Medicare program. Thus, DME supplier do provide services such as fitting and adjusting DME items. For example, wheelchair training. Generally DME personnel perform some therapeutic services but not generally diagnostic (Bach, 2009). Since DME suppliers deal with tangible items, the typical payment process is a fee schedule approach, a given piece of DME will be reimbursed at a given fee schedule amount. The fee schedule amounts may include rental and used equipment. Thus, no intensity level of care provider needs low level of ability to treat patients, and relatively simple payment process.

Low intensity care provider Low intensity level of care provider can provide ambulatory care services. Typically, clinics belong in this category. This provider group provides primary care services, community services, and outpatient clinic services. We can call this service as office-based medical service, and this area has been widely explored. Cayirli and Veral (2003) and Gupta and Denton (2008) provide extensive literature of clinic services. On the other hand, a healthcare provider, who delivers home care services, also includes this category. Home care providers deliver medical treatment using resources of professional nurses, home health aides, medical equipment, and other supporting items (Hulshof et al. 2012). Home health care providers primarily treat the elderly, disabled, and chronically ill, but not life-threatened ill (Hans et al., 2012). Since primary care and home health care does not require a high level of capability to deal with severe patients, low intensity level care needs low providers treatment capability, and corresponding low payment procedure.

Moderate intensity care provider Medium intensity level of care providers can provide care to patients whose degree of severity is high enough to be inpatient. General hospitals are usually located in this category. They provide both surgical and medical care services, and these surgical services are usually associated with a length of stay (Guerriero and Guido 2011). Care providers who focus on residential services can be classified into this group. Residential services usually cover elderly patients, who need a higher level of treatment than home care service, but do not strictly have to be in a hospital (Hare et al., 2009). Thus, this type of care provider requires a higher level of care capability, and their payment process is more complex than the low or no intensity levels of care provider.

High intensity care provider Finally, the highest intensity level of care can pertains to major surgical or medical intervention, which are directly associated with a fatal illness, such as cancer or myocardial infarction. Some special hospitals, which have an academic research institute, are categorized in this group. If medium level of care providers cannot resolve a patients illness, the care provider sends patients to the high intensity care provider. Since this type of care providers is capable to deal with the highest severity of illness, they design excellent emergency care services, such as rapid response of an ambulance to the health care provider emergency center (Green and Kolesar, 2004). In addition, this provider group sometimes will provide rapid response to tackle a potentially fatal, acute or chronic condition or illness, they need the highest level of treatment capability, which may involve a wide range of procedures and associated facilities. In addition, the primary goal of this provider is to reduce mortality, and may require many possible alternatives to attain this goal. This problem results in third-party payers to adopt the most complex payment processes.

APPENDIX C

PROOF OF THEOREMS AND LEMMAS IN SECTION 3

Proof of Lemma 1:

If Part: When $t = 1$, we have $b_1 = \mu_1 \geq L$. Thus, we have $d_1 = (L - b_1)^+ = 0$.

When $t = 2$, we have $b_2 = (b_1 - L)^+ + \mu_2 = b_1 - L + \mu_2 = \mu_1 + \mu_2 - L$. Since $\sum_{i=1}^{2} \mu_i \geq 2L$, we have $d_2 = (L - b_2)^+ = (2L - \sum_{i=1}^{2} \mu_i)^+ = 0$.

Assume the property holds for Period $t-1$, i.e., $b_{t-1} = \sum_{i=1}^{t-1} \mu_i - (t-2)L$, $d_{t-1} = (L - b_{t-1})^+ = ((t-1)L - \sum_{i=1}^{t-1} \mu_i)^+ = 0$. Then, in Period t, we have $b_t = (b_{t-1} - L)^+ + \mu_t = \sum_{i=1}^{t} \mu_i - (t-1)L$. Since $\sum_{i=1}^{t} \mu_i \geq tL$, we have $d_t = (L - b_t)^+ = (tL - \sum_{i=1}^{t} \mu_i)^+ = 0$. Thus, $D(\pi) = \sum_{i=1}^{r} d_i = 0$.

Since $rL = \sum_{i=1}^{r} \mu_i$, we have $O(\pi) = \sum_{i=1}^{r} \mu_i + D(\pi) - rL = 0$.

Only If Part: Assume there exists Period t, $1 \leq t \leq r$, which is the earliest period that $\sum_{i=1}^{t} \mu_i < tL$. Thus, the physician is idle after serving the first t patients, which contradicts the statement that physician's idle time is 0. Thus, if A block schedule π has zero physician's idle time and overtime, we have $\sum_{i=1}^{t} \mu_i \geq tL$, $\forall 1 \leq t \leq r$. This completes the proof. ∎

The Schedule s_{ub} in Set Λ which Provides the Upper Bound of Patient Waiting Time:

We define block schedule $\pi_{ub} = (A, A, ..., B, B, B, ...)$ as follows: r_a Type A patients are scheduled consecutively followed by r_b Type B patients in the block i, $i = 1, 2, \ldots, k$. This block π_{ub} is repeated k times to cover T time slots to form s_{ub}, where $s_{ub} = (\pi_{ub}, \pi_{ub}, \ldots, \pi_{ub})$. In our example, since $r_a = 2$, $r_b = 3$, we have

$\pi_{ub} = (A, A, B, B, B)$,

$s_{ub} = (A, A, B, B, B, A, A, B, B, B,\ A, A, B, B, B, A, A, B, B, B, A, A, B, B, B)$. In Lemma 1.1, we prove that Schedule s_{ub} is in Set Λ. Lemma 1.2 describes the property regarding patients' waiting time.

Lemma 1.1: Schedule s_{ub} is in Set Λ.

Proof of Lemma 1.1: According to Lemma 1, to prove Schedule $s_{ub} \in \Lambda$, we need to show $\sum_{i=1}^{t} \mu_i \geq tL$, $1 \leq t \leq r$.

In π_{ub}, the first r_a patients are of Type A. Since $\mu_a \geq L$, it is easy to see that $\sum_{i=1}^{t} \mu_i \geq tL$, $1 \leq t \leq r_a$.

In π_{ub}, Type B patients are scheduled in Period $r_a + j$, $1 \leq j \leq r_b$. Thus, for $t = r_a + j$, $1 \leq j \leq r_b$, we have:

$$\begin{aligned}
\sum_{i=1}^{t} \mu_i - tL &= r_a\mu_a + j\mu_b - (r_a + j)L \\
&= r_a(\mu_a - L) - j(L - \mu_b) \\
&= r_a(\mu_a - \frac{r_a\mu_a + r_b\mu_b}{r_a + r_b}) - j(\frac{r_a\mu_a + r_b\mu_b}{r_a + r_b} - \mu_b) \\
&= \frac{r_a r_b(\mu_a - \mu_b)}{r_a + r_b} - j\frac{r_a(\mu_a - \mu_b)}{r_a + r_b} \\
&= \frac{r_a(\mu_a - \mu_b)}{r_a + r_b}(r_b - j) \geq 0
\end{aligned}$$

The result follows. ■

Thus, Schedule s_{ub} is in Set Λ. Next, in Lemma 1.2, we calculate the patients' waiting time in Schedule s_{ub}, and prove that this value represents the upper bound of the patients' waiting time of any schedule $s \in \Lambda$.

Lemma 1.2: The patients' waiting time in Schedule s_{ub} is $W(s_{ub}) = k[\frac{r_a(r_a+2r_b-1)}{2}(\mu_a - L) - \frac{r_b(r_b-1)}{2}(L - \mu_b)]$, which represents the upper bound of the patients' waiting time

$W(s)$ of any schedule $s \in \Lambda$, i.e., $W(s) \leq W(s_{ub})$.

Proof of Lemma 1.2: The patients' waiting time of π_{ub} can be derived as follows:

$$\begin{aligned}
W(\pi_{ub}) &= \sum_{j=1}^{r_a}(j-1)(\mu_a - L) + \sum_{j=1}^{r_b}[r_a\mu_a + (j-1)\mu_b - (r_a + j - 1)L] \\
&= \sum_{j=1}^{r_a}(j-1)(\mu_a - L) + r_a r_b(\mu_a - L) + \sum_{j=1}^{r_b}(j-1)(\mu_b - L) \\
&= \frac{r_a(r_a-1)}{2}(\mu_a - L) + r_a r_b(\mu_a - L) - \frac{r_b(r_b-1)}{2}(L - \mu_b) \\
&= \frac{r_a(r_a + 2r_b - 1)}{2}(\mu_a - L) - \frac{r_b(r_b-1)}{2}(L - \mu_b)
\end{aligned}$$

Since in block π_{ub} the last patient leaves at $(r_a + r_b)L$, the first patient in the next block does not wait for service. Similarly, the first patient in each block does not wait for service. Thus:

$$W(s_{ub}) = kW(\pi_{ub}) = k[\frac{r_a(r_a + 2r_b - 1)}{2}(\mu_a - L) - \frac{r_b(r_b-1)}{2}(L - \mu_b)].$$

Suppose schedule $s \in \Lambda$, $s \neq s_{ub}$, provides the maximum total patient waiting time. Note that in s_{ub}, all patients of Type B are scheduled at the end of the block after serving all patients of Type A. Thus, in s, there must exist at least a patient of Type B who is scheduled before Type A patients. Let k be the first period in which a Type B patient is scheduled before Type A patients (see Figure C.1). Since there are r_a Type A patients, and $s \neq s_{ub}$, we have $1 \leq k \leq r_a$. We now perform the following operation on s: move the k^{th} patient (Type B) to the end of schedule s, i.e., to the r^{th} position in s. We denote the new block schedule as s' (see Figure C.1). Thus, we have

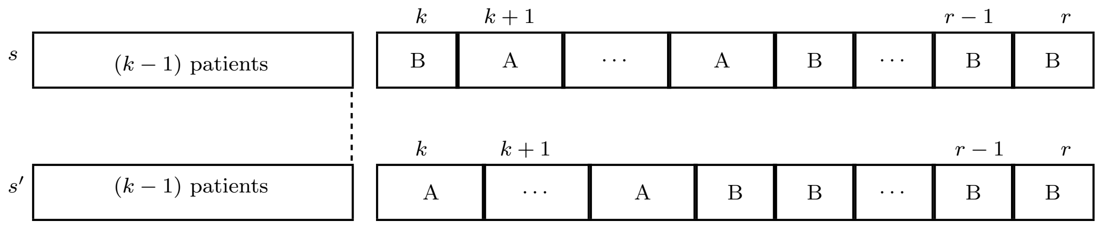


Figure C.1: Block schedules s and s' when $r = r_a + r_b$.

$$\mu_i = \mu'_i = \mu_a, 1 \leq i \leq k-1; \mu_k = \mu_b;$$

$$\mu'_{k-1+i} = \mu_{k+i}, 1 \leq i \leq r-k; \mu'_r = \mu_b.$$

First we show that Schedule s' is also in Set Λ. Since Schedule s is in Set Λ, we have $\sum_{i=1}^{t} \mu_i \geq tL$, $1 \leq t \leq r$. Thus, for $t = k+j$, $1 \leq j \leq r-k$, we have $\sum_{i=1}^{k-1} \mu_i + \mu_k + \sum_{i=1}^{j} \mu_{k+i} - (k+j)L = \sum_{i=1}^{k} \mu_i + \mu_b + \sum_{i=1}^{j} \mu_{k+i} - (k+j)L \geq 0$. Thus, $\sum_{i=1}^{k} \mu_i + \sum_{i=1}^{j} \mu_{k+i} - (k+j-1)L \geq L - \mu_b > 0$.

In Schedule s', for $1 \leq t \leq k-1$, we have $\sum_{i=1}^{t} \mu'_i = t\mu_a \geq tL$. For $t = k+j-1$, $1 \leq j \leq r-k$, we have $\sum_{i=1}^{t} \mu'_i - tL = \sum_{i=1}^{k-1} \mu'_i + \sum_{i=1}^{j} \mu'_{k+i-1} - (k+j-1)L$. Since $\mu_i = \mu'_i = \mu_a, 1 \leq i \leq k-1; \mu'_{k-1+i} = \mu_{k+i}, 1 \leq i \leq r-k$, we have $\sum_{i=1}^{k-1} \mu'_i + \sum_{i=1}^{j} \mu'_{k+i-1} - (k+j-1)L = \sum_{i=1}^{k-1} \mu_i + \sum_{i=1}^{j} \mu_{k+i} - (k+j-1)L \geq 0$. For $t = r$, we have $\sum_{i=1}^{t} \mu'_i = tL$. Thus, Schedule s' is also in Set Λ.

Since Schedules s and s' are in Set Λ, the physician's idle time and overtime are 0. Thus, we have

$$w_1 = w'_1 = 0; \quad w_t = \sum_{i=1}^{t-1} \mu_i - (t-1)L, \quad w'_t = \sum_{i=1}^{t-1} \mu'_i - (t-1)L, \quad 2 \leq t \leq r.$$

Let $W(s)$ (respectively, $W(s')$) be the total waiting time of the patients scheduled in s (respectively, s'). Thus, we have $W(s) = k\sum_{t=1}^{r} w_t$, $W(s') = k\sum_{t=1}^{r} w'_t$. Then,

$$\begin{aligned}
\sum_{t=1}^{r} w'_t - \sum_{t=1}^{r} w_t &= \sum_{t=1}^{k}(w'_t - w_t) - w_{k+1} + \sum_{j=1}^{r-k-1}(w'_{k+j} - w_{k+j+1}) + w'_r \\
&= 0 - [(k-1)\mu_a + \mu_b - kL] + \\
&\quad \sum_{j=1}^{r-k-1}(L + \sum_{i=1}^{k+j-1}\mu'_i - \sum_{i=1}^{k+j}\mu_i) + (L - \mu_b) \\
&= kL - (k-1)\mu_a - \mu_b + (r-k-1)(L-\mu_b) + (L-\mu_b) \\
&= rL - (k-1)\mu_a - (r-k+1)\mu_b \\
&= (r_a + 1 - k)(\mu_a - \mu_b).
\end{aligned}$$

Since $k \leq r_a$, $\mu_a > \mu_b$, we have $\sum_{t=1}^{r} w'_t - \sum_{t=1}^{r} w_t > 0$. Thus, $W(s') > W(s)$, which contradicts the statement that schedule $s \in \Lambda$, $s \neq s_{ub}$, provides the maximum total patient waiting time. Hence the lemma is proved. ∎

Proof of Lemma 2:

The while loop performs at most $r_a + r_b$ iterations. Thus, the complexity of Algorithm *OptBlock(π_o)* is $O(r_a + r_b)$. ∎

Proof of Lemma 3:

We first show that Algorithm OptBlock(π_o) finds a feasible schedule. Note that the first patient scheduled in π_o is Type A. Since $\mu_a > L = \frac{r_a\mu_a + r_b\mu_b}{r_a + r_b}$, there is no idle time between patients scheduled in periods 1 and 2. Similarly, Step 0 assures that no idle time occurs between patients scheduled in periods $j-1$ and j, for $j = 3, 4, \ldots, r_a + r_b$. According to Lemma 1, π_o has zero physician's idle time and overtime. Thus, $s_o = (\pi_o, \pi_o, \ldots, \pi_o)$ is a feasible solution of Problem SP_0. Next, we show that Algorithm *OptBlock(π_o)* finds an optimal schedule for Problem SP_0.

Suppose schedule s_u ($s_u \neq s_0$) provides the minimum patients' waiting time. Let $s_u = (\pi_u, \pi_u, \ldots, \pi_u)$. Therefore, there must exist periods j and ℓ in which different types of patients are scheduled in π_u and π_o. Let $j < \ell$ and (j, ℓ) be the smallest indices in which this difference occurs in π_u and π_o. Note that the schedule of the first $(j-1)$ patients are identical in both π_u and π_o. We now consider two cases:

Case 1: In period j, Type B is scheduled in π_u and Type A is scheduled in π_o. In period ℓ, Type A is scheduled in π_u and Type B is scheduled in π_o. According to our algorithm, a Type A patient can be scheduled in period j only if $F_{j-1} + \mu_b < jL$. Therefore, the physician's idle time is positive in period j of π_u. This contradicts our assumption that the schedule π_u is feasible.

Case 2: In period j, Type A is scheduled in π_u and Type B is scheduled in π_o. In period ℓ, Type B is scheduled in π_u and Type A is scheduled in π_o.

We now perform the following operation on π_u: interchange patients in periods j and ℓ. Call the new schedule π'_u. Note that the schedule of the first ℓ patients are identical in both π'_u and π_o. Therefore, the new schedule π'_u is feasible.

The proof is by contradiction. Since both π_u and π'_u are feasible, there is no idle time in those schedules. According to our assumption, π_u provides the minimum patients' waiting time, i.e., the patients' waiting time $W_u \leq W'_u$. The schedule of the first $(j-1)$ patients and the last $(r_a + r_b - \ell)$ patients are identical in both π_u and π'_u. Consequently, the start times of the first j patients and the last $(r_a + r_b - \ell)$ patients are the same in both π_u and π'_u. Thus, the total waiting time of the first j patients and the last $(r_a + r_b - \ell)$ patients are the same in both π_u and π'_u. Note that the waiting time of the $(j+1)^{th}$ patient is $w_{j+1} = F_{j-1} + \mu_a - jL$ (respectively,

$w'_{j+1} = F_{j-1} + \mu_b - jL$) in π_u (respectively, π'_u). Therefore, we have $w_{j+1} - w'_{j+1} = \mu_a - \mu_b > 0$. Similarly, we have $w_i - w'_i = \mu_a - \mu_b$, for $i = j+1, j+2, \ldots, \ell$. Thus, $W_u - W'_u = (\ell - j)(\mu_a - \mu_b) > 0$. This implies $W_u > W'_u$, which contradicts the claim that π_u provides the minimum patients' waiting time. ∎

Proof of Observation 1:

We illustrate this in the following counter example: Assume there are three types of patients: $\mu_1 = 8$, $\mu_2 = 11$, $\mu_3 = 13$; $r_1 = 2, r_2 = 1, r_3 = 1$. Thus, L= (2*8+11*1+13*1)/(2+1+1)=10.

If we use $OptBlock_m(\pi)$, then we have π= {Type 2, Type 3, Type 1, Type 1}. The patients' waiting time = 0 + 1 + 4 + 2 = 7.

However, the optimal schedule π^* should be {Type 3, Type 1, Type 2, Type 1} with the patients' waiting time = 0 + 3 + 1 + 2 = 6. ∎

Proof of Lemma 4:

When we implement DP_m for each period, at most m possible types of patients can be assigned. Since there are r periods in one block, at most m^r possible scheduling sequences can be considered. Thus, the complexity of DP_m is $O(m^r)$. ∎

Proof of Theorem 1:

Consider an arbitrary instance of a known NP-complete problem: Numerical Matching with Target Sum (NMTS) (Garey and Johnson 1979).

Numerical Matching with Target Sums (NMTS): Given three sets of positive integers $S_x = \{x_1, \ldots, x_n\}$, $S_y = \{y_1, \ldots, y_n\}, S_z = \{z_1, \ldots, z_n\}$, can $S_y \cup S_z$ be partitioned into n disjoint subsets $\Gamma_1, \ldots, \Gamma_n$ with $\Gamma_k = \{y_{i_k}, z_{j_k}\}$ such that $x_k = y_{i_k} + z_{j_k}$, for $k = 1, \ldots, n$?

Given an instance of NMTS, we construct a specific instance of the decision version of problem SP_0^m as follows: We assume that $\Sigma_x = \Sigma_y + \Sigma_z$, where $\Sigma_x = \sum_{i=1}^n x_i$, $\Sigma_y = \sum_{i=1}^n y_i$ and $\Sigma_z = \sum_{i=1}^n z_i$. There are $m = 3n$ patient types P_ℓ, $\ell = 1, 2, \ldots, m$, and $r_\ell = 1$ for all $\ell = 1, 2, \ldots, m$.

Let $B = (n + 1)\Sigma_x$, $M = (n + 2)B$. For convenience, the m patient types are classified into three classes of patients as follows:

i. Class X patients P_i^x with $\mu_i^x = 3M - B - x_i$, $i = 1, 2, \ldots, n$.

ii. Class Y patients P_i^y with $\mu_i^y = y_i$, $i = 1, 2, \ldots, n$.

iii. Class Z patients P_i^z with $\mu_i^z = B + z_i$, $i = 1, 2, \ldots, n$.

Thus, the length of each period, $L = \frac{\sum_{i=1}^n \mu_i^x + \sum_{i=1}^n \mu_i^y + \sum_{i=1}^n \mu_i^z}{3n}$ $= \frac{3nM - nB - \sum_{i=1}^n x_i + \sum_{i=1}^n y_i + nB + \sum_{i=1}^n z_i}{3n} = M$. A threshold value $D = 3nM - 2nB - 2\Sigma_x + \Sigma_y$.

<u>Decision Problem (DQ)</u>: Does there exist a schedule of patients σ such that the physician idle time is zero, and the total patient waiting time, W_σ satisfies $W_\sigma \leq D = 3nM - 2nB - 2\Sigma_x + \Sigma_y$?

The decision problem is clearly in class NP. Also, it is easy to verify that the construction of the decision problem can be done in polynomial time. We now show that there exists a schedule σ such that $W_\sigma \leq 3nM - 2nB - 2\Sigma_x + \Sigma_y$ if and only if there exists a solution to the NMTS problem.

<u>*If Part*</u>: Suppose there exists a NMTS partition. Without loss of generality, we may assume $x_i = y_i + z_i$, $i = 1, 2, \ldots, n$. We denote patient schedule σ as $(\sigma(1), \sigma(2), \ldots, \sigma(m))$, where $\sigma(i)$ denotes the i^{th} patient scheduled in schedule σ. Let s_i and f_i be the start time and finish time of the patient scheduled in the i^{th} period. Let w_i be the waiting time of the i^{th} patient. Consider the following schedule

$\sigma = (P_1^x, P_1^y, P_1^z, P_2^x, P_2^y, P_2^z, \ldots, P_n^x, P_n^y, P_n^z)$ shown in Table C.1:

Table C.1: Patient Schedule σ with $W_\sigma = 3nM - 2nB - 2\Sigma_x + \Sigma_y$.

i	1	2	3	4	5	6	...	$3n-2$	$3n-1$	$3n$
Type	P_1^x	P_1^y	P_1^z	P_2^x	P_2^y	P_2^z	...	P_n^x	P_n^y	P_n^z
$\mu_{\sigma(i)}$	$3M-B$ $-x_1$	y_1	$B+z_1$	$3M-B$ $-x_2$	y_2	$B+z_2$	... ...	$3M-B$ $-x_n$	y_n	$B+z_n$
s_i	0	$3M-B$ $-x_1$	$3M-B$ $-x_1+y_1$	$3M$	$6M-B$ $-x_2$	$6M-B$ $-x_2+y_2$	... ...	$3(n-1)M$	$3nM-B$ $-x_n$	$3nM-B$ $-x_n+y_n$
f_i	$3M-B$ $-x_1$	$3M-B$ $-x_1+y_1$	$3M$	$6M-B$ $-x_2$	$6M-B$ $-x_2+y_2$	$6M$	... ...	$3nM-B$ $-x_n$	$3nM-B$ $-x_n+y_n$	$3nM$
w_i	0	$2M-B$ $-x_1$	$M-B$ $-x_1+y_1$	0	$2M-B$ $-x_2$	$M-B$ $-x_2+y_2$	... ...	0	$2M-B$ $-x_n$	$M-B$ $-x_n+y_n$

Now it is easy to see that $W_\sigma = \sum_{i=1}^{3n} w_i = 3nM - 2nB - 2\Sigma_x + \Sigma_y = D$.

Only If Part: Suppose there exists a schedule σ^* such that $W_{\sigma^*} \leq D$ with zero physician idle time. First we show that σ^* is similar to that shown in Table C.1. Since $M = (n+2)B$, $B = (n+1)\Sigma_x$, we have $\mu_i^x > M > \mu_i^z > B > \mu_i^y$, for $i = 1, 2, \ldots, n$.

Claim 1: In schedule σ^*, the patient assigned to Period 1, $\sigma(1)$, must be a Class X patient.

Proof of Claim 1: To guarantee the physician idle time in Period 1 is 0, we must have $\mu_{\sigma(1)} \geq L = M$. Note that $\mu_i^x > M$, $\mu_i^y < M$, and $\mu_i^z < M$, for $i = 1, 2, \ldots, n$. Thus, $\sigma(1)$ must be a Class X patient. □

Since there are n Class X patients in schedule σ^*, we divide σ^* into n blocks such that (i) there is only one Class X patient in each block, and (ii) each block begins with a Class X patient. Without loss of generality, we may assume the Class X patient assigned in Block i is P_i^X, $i = 1, 2, \ldots, n$. We denote $x_{max} = \max_{i=1}^n \{x_i\}$, $z_{max} = \max_{i=1}^n \{z_i\}$.

Claim 2: In schedule σ^*, if there is exactly one Non-Class-X patient in Block i,

$1 \le i \le n$, then the waiting time of this Non-Class-X patient is at least $2M - B - x_{max}$.

Proof of Claim 2: Assume the Class X patient in Block i (the i^{th} Class X patient) is in Period t, $1 \le t < n$. Thus, the Non-Class-X patient in Block i is in Period $t+1$. The waiting time of this Class X patient is at least 0, i.e., $w_t \ge 0$. The service time of the i^{th} Class X patient, $\mu_{\sigma(t)}$ is at least $3M - B - x_{max}$, i.e., $3M - B - x_i \ge 3M - B - x_{max}$. Since the length of each time slot $L = M$, then $w_{t+1} = (w_t + \mu_{\sigma(t)} - L)^+ \ge (2M - B - x_{max})^+$. Since $M = (n+2)B$, $B = (n+1)\Sigma_x$, we have $(2M - B - x_{max})^+ = 2M - B - x_{max}$. Thus, the waiting time of the Non-Class-X patient who is scheduled in Block i is at least $2M - B - x_{max}$. □

Claim 3: In σ^*, for $k \ge 2$, if there are exactly k Non-Class-X patients in Block i, $1 \le i \le n$, then the average waiting time of these k Non-Class-X patients is at least $\frac{k+1}{2}(M - B - z_{max})$.

Proof of Claim 3: In Block i, we first consider the k^{th} Non-Class-X patient. Assume this patient is in Period t, $k \le t \le n$. Since there is no physician's idle time in Period t, we must have the sum of the waiting time and the service time of this patient is at least M, i.e., $w_t + \mu_{\sigma(t)} \ge M$. Since the maximum service time of all Non-Class-X patients is $B + z_{max}$, i.e., $\mu_{\sigma(t)} \le B + z_{max}$. Thus, we have $w_t \ge M - B - z_{max}$. Also since $w_t = w_{t-1} + \mu_{\sigma(t-1)} - M$, we have $w_{t-1} = w_t + M - \mu_{\sigma(t-1)} \ge 2(M - B - z_{max})$. Similarly, the waiting time of the first Non-Class-X patient, w_{t+1-k} is at least $k(M - B - z_{max})$. Thus, we have the total waiting time of these k Non-Class-X patients, $\sum_{i=1}^{k} w_{t+1-k} \ge \sum_{i=1}^{k} i(M - B - z_{max}) = \frac{k(k+1)}{2}(M - B - z_{max})$. Then, the average waiting time of these k Non-Class-X patients is at least $\frac{k+1}{2}(M - B - z_{max})$.

□

Claim 4: In σ^*, if there are exactly k Non-Class-X patients in Block i, $k = 1$ or $k \ge 3$, $1 \le i \le n$, then the average waiting time of these k Non-Class-X patients is at least $2(M - B - z_{max})$.

Proof of Claim 4: If $k \geq 3$, according to Claim 3, we have the average waiting time of these k Non-Class-X patients is at least $\frac{k+1}{2}(M - B - z_{max})$. Since $\frac{k+1}{2}(M - B - z_{max})$ is increasing with an increase in k, we have $\frac{k+1}{2}(M - B - z_{max}) \geq \frac{3+1}{2}(M - B - z_{max}) = 2(M - B - z_{max})$, for $k \geq 3$.

If $k = 1$, according to Claim 2, we have the waiting time of this Non-Class-X patient is at least $2M - B - x_{max}$. We have $2M - B - x_{max} - 2(M - B - z_{max}) = B - x_{max} + 2z_{max}$. Since $B = (n+1)\Sigma_x$, then $B - x_{max} + 2z_{max} > 0$. □

Claim 5: In σ^*, there are exactly 2 Non-Class-X patients in each Block i, $1 \leq i \leq n$.

Proof of Claim 5: If not all blocks have exactly 2 Non-Class-X patients, then we can have at most $n - 2$ blocks with exactly 2 Non-Class-X patients. Assume there are n_1 blocks with exactly 2 Non-Class-X patients, $0 \leq n_1 \leq n - 2$. Thus, according to Claim 3, these n_1 blocks include $2n_1$ Non-Class-X patients with average waiting time at least $\frac{3}{2}(M - B - z_{max})$. The remaining $2(n - n_1)$ Non-Class-X patients are in blocks with either 1 or at least 3 Non-Class-X patients.

Since the waiting time of Class-X patient is non-negative, then the total waiting time of all patients should be at least the total waiting time of all Non-Class-X patients. Thus, we have $W_{\sigma^*} \geq 2n_1 * \frac{3}{2}(M - B - z_{max}) + 2(n - n_1) * 2(M - B - z_{max}) = (4n - n_1)(M - B - z_{max})$. Then $(4n - n_1)(M - B - z_{max}) - D = (n - n_1)M - (2n - n_1)B + 2\Sigma_x - \Sigma_y - (4n - n_1)z_{max}$. Since $0 \leq n_1 \leq n - 2$, $\Sigma_x > \Sigma_y$, $\Sigma_x > z_{max}$, we have $(n - n_1)M - (2n - n_1)B + 2\Sigma_x - \Sigma_y - (4n - n_1)z_{max} \geq 2M - 2nB + (1 - 4n)\Sigma_x$. Since $M = (n+2)B$, $B = (n+1)\Sigma_x$, we have $2M - 2nB + (1 - 4n)\Sigma_x = 5\Sigma_x > 0$. This contradicts the assumption that $W_{\sigma^*} \leq D$. Thus, all n blocks must have exactly 2 Non-Class-X patients. ■

According to Claims 1 and 5, the patient assigned to Period $3i-2$ is Class X, i.e., $\sigma(3i-2) = P_i^x$, $i = 1, 2, \ldots, n$. Next, we show that in Block k, $1 \leq k \leq n$, the sequence of 2 Non-Class-X patients is $(P_{i_k}^y, P_{j_k}^z)$. We denote $y_{max} = \max_{i=1}^n\{y_i\}$, $z_{min} = \min_{i=1}^n\{z_i\}$.

Claim 6: In Block k ($1 \leq k \leq n$) of σ^*, if the two Non-Class-X patients are not in the sequence: $(P_{i_k}^y, P_{j_k}^z)$, then the total waiting time of these two Non-Class-X patients is at least $(3M - B - 2x_k + z_{min})$.

Proof of Claim 6: Since the sequence is not $(P_{i_k}^y, P_{j_k}^z)$, there are two possible scenarios:

1. $(P_{i_k}^y, P_{j_k}^y)$

   We first consider the second Class-Y patient. Since there is no idle time, the sum of the waiting time and the service time of this patient is at least M. Since the maximum service time of all Class-Y patients is y_{max}. The second Class-Y patient's waiting time is at least $M - y_{max}$. Similarly, the waiting time of the first Class-Y patient is at least $2(M - y_{max})$. Thus, the total waiting time of these two Non-Class-X patients is at least $3(M - y_{max})$. Since $B = (n+1)\Sigma_x$, we have $3(M - y_{max}) - (3M - B - 2x_k + z_{min}) = B + 2x_k - 3y_{max} - z_{min} > 0$.

2. $(P_{i_k}^z, P_{j_k}^y)$ or $(P_{i_k}^z, P_{j_k}^z)$

   The waiting time of the Class X patient in Block k is at least 0, and the service time of this Class X patient is $3M - B - x_k$. Since the length of each time slot $L = M$, then the waiting time of the first Class Z patient is at least $2M - B - x_k$. Since the minimum service time of this Class Z patient is $B + z_{min}$, then the waiting time of the patient who follows the first Class Z patient is at least $M - x_k + z_{min}$. Thus, the total waiting time of these two Non-Class-X patients is at least $3M - B - 2x_k + z_{min}$.

Thus, under both scenarios, we have that the total waiting time of these two Non-Class-X patients is at least $(3M - B - 2x_k + z_{min})$. □

Claim 7: In each block k ($1 \leq k \leq n$) of σ^*, the two Non-Class-X patients are in the sequence: $(P^y_{i_k}, P^z_{j_k})$.

Proof of Claim 7: Assume there are n_2 blocks which don't follow the sequence: $(P^y_{i_k}, P^z_{j_k})$, $1 \leq n_2 \leq n$. Thus, the remaining $n - n_2$ blocks follows the sequence.

If Block k follows the sequence $(P^y_{i_k}, P^z_{j_k})$, then the waiting time of the Class Y patient is at least $2M - B - x_k$. Since the minimum service time of the Class Y patient is y_{min}, then the waiting time of the Class Z patient is at least $M - B - x_k + y_{min}$. Thus, the total waiting time of the two Non-Class-X patients in Block k is at least $3M - 2B - 2x_k + y_{min}$.

Since the waiting time of Class-X patient is non-negative, then the total waiting time of all patients should be at least the total waiting time of all Non-Class-X patients. Thus, we have $W_{\sigma^*} \geq -2\sum_{k=1}^{n} x_k + (n - n_2)(3M - 2B + y_{min}) + n_2(3M - B + z_{min}) = -2\Sigma_x + 3nM + (n_2 - 2n)B + (n - n_2)y_{min} + n_2 z_{min}$. We have $-2\Sigma_x + 3nM + (n_2 - 2n)B + (n - n_2)y_{min} + n_2 z_{min} - D = n_2 B + (n - n_2)y_{min} + n_2 z_{min} - \Sigma_y$. Since $n_2 \geq 1$, $B = (n+1)\Sigma_x$, we have $n_2 B + (n - n_2)y_{min} + n_2 z_{min} - \Sigma_y > 0$, which contradicts the assumption that $W_\sigma \leq D$. This completes the proof. ■

Thus, in each block k ($1 \leq k \leq n$) of σ^*, the three patients are in the sequence: $(P^x_k, P^y_{i_k}, P^z_{j_k})$. Next we show that given a schedule of patients σ such that the physician idle time is zero, and $W_\sigma \leq D = 3nM - 2nB - 2\Sigma_x + \Sigma_y$, we find a solution to NMTS.

Claim 8: In Block k of σ^*, the waiting time of the Class X patient is 0, i.e.,

$w_{\sigma(3k-2)} = 0$, $1 \leq k \leq n$.

Proof of Claim 8: According to Claim 7, in Block k of σ^*, the three patients are in the sequence: $(P_k^x, P_{i_k}^y, P_{j_k}^z)$. Thus, the waiting time of these three patients are $w_{\sigma(3k-2)}$, $w_{\sigma(3k-2)} + 2M - B - x_k$, and $w_{\sigma(3k-2)} + M - B - x_k + y_{i_k}$. Then the total waiting time of these three patients is $3w_{\sigma(3k-2)} + 3M - 2B - 2x_k + y_{i_k}$. Therefore $W_{\sigma^*} = \sum_{k=1}^n (3w_{\sigma(3k-2)} + 3M - 2B - 2x_k + y_{i_k}) = 3\sum_{k=1}^n w_{\sigma(3k-2)} + 3nM - 2nB - 2\Sigma_x + \Sigma_y = 3\sum_{k=1}^n w_{\sigma(3k-2)} + D$. Since $W_{\sigma^*} \leq D$, we have $\sum_{k=1}^n w_{\sigma(3k-2)} = 0$. Thus, $w_{\sigma(3k-2)} = 0$, $1 \leq k \leq n$. □

Note that $w_{\sigma(4)} = w_{\sigma(1)} + \mu_{\sigma(1)} + \mu_{\sigma(2)} + \mu_{\sigma(3)} - 3L = w_{\sigma(1)} - x_1 + y_{i_1} + z_{j_1}$. Since $w_{\sigma(1)} = w_{\sigma(4)} = 0$, we have $-x_1 + y_{i_1} + z_{j_1} = 0$. Similarly, we can show that $-x_k + y_{i_k} + z_{j_k} = 0$ for $2 \leq k \leq n-1$. Since $\Sigma_x = \Sigma_y + \Sigma_z$, we also have $-x_n + y_{i_n} + z_{j_n} = 0$. Thus, we obtain a solution to NMTS. This completes the proof of **Theorem 1**. ■

Proof of Theorem 2:

First, we describe which two scenarios to compare. Given a vector of $(z_1, z_2, \ldots, z_r)$, we interpret it in the following way: z_1 represents the status of the first Type A patient, z_2 represents the status of the second Type A patient, …, z_{r_a} represents the status of the r_a^{th} Type A patient; z_{r_a+1} represents the status of the first Type B patient, z_{r_a+2} represents the status of the second Type B patient, , …, z_r represents the status of the r_b^{th} (last) Type B patient. For any possible vector of $(z_1, z_2, \ldots, z_r)$, $z_i \in \{0,1\}, 1 \leq i \leq r$, we can find one and only one corresponding scenario in each sequence.

Independent of the sequence, for the corresponding scenario, we have $x_a = \sum_{i=1}^{r_a} z_i$, $x_b = \sum_{i=r_a+1}^{r} z_i$. Among the scheduled r_a (respectively, r_b) Type A (respectively, Type B) patients, x_a (respectively, x_b) show up, while $(r_a - x_a)$ (respectively,

$(r_b - x_b)$) do not show up. Thus, for each scenario, the probability has the same value: $(1-p_a)^{x_a} p_a^{r_a - x_a} (1-p_b)^{x_b} p_b^{r_b - x_b}$.

Next, we calculate the physician's idle time in each scenario. If a sequence satisfies Lemma 1, then if all scheduled patients show up, the physician works for rL, with 0 idle time. If only x_a Type A and x_b Type B patients show up, the physician needs to treat all of these patients. Thus, the physician's working time is $x_a\mu_a + x_b\mu_b$. Therefore, the physician's idle time is $(rL - x_a\mu_a - x_b\mu_b)$. Note that this number is independent of the specific sequence. To calculate the physician's expected idle time, for each scenario, we need to multiply the probability of that scenario. The probability $(1-p_a)^{x_a} p_a^{r_a - x_a} (1-p_b)^{x_b} p_b^{r_b - x_b}$ is also independent of the specific sequence.

Thus, for any one of the 2^r possible vectors of $(z_1, z_2, \ldots, z_r)$, $z_i \in \{0, 1\}, 1 \leq i \leq r$, we can find one and only one corresponding scenario in each sequence. Each scenario has the same value in probability and physician's idle time. Therefore, for any schedule in Set Λ, the expected physician's idle time has the same value. ■

Proof of Theorem 3:

Assume π_1 is an arbitrary sequence satisfying Lemma 1, π_2 is an arbitrary sequence not satisfying Lemma 1. Given a vector of $(z_1, z_2, \ldots, z_r)$, we interpret it in the following way: z_1 represents the status of the first Type A patient, z_2 represents the status of the second Type A patient, ..., z_{r_a} represents the status of the r_a^{th} Type A patient; z_{r_a+1} represents the status of the first Type B patient, z_{r_a+2} represents the status of the second Type B patient, , ..., z_r represents the status of the r_b^{th} (last) Type B patient. For any possible vector of $(z_1, z_2, \ldots, z_r)$, $z_i \in \{0, 1\}, 1 \leq i \leq r$, we can find one corresponding scenario s_1 in sequence π_1, one corresponding scenario s_2 in sequence π_2.

First, consider $z = (1,1,1,1,1)$, when all patients showed up. According to Lemma 1, in Scenario s_1, the physician's working time, T_1 is rL. Meanwhile, in Scenario s_2, the physician's working time, T_2, is more than rL (i.e., $T_2 > rL$) with a positive amount of overtime.

Then, for an arbitrary z vector, some patients may not show up. To treat x_a Type A patients and x_b Type B patients, the physician needs to spend $(x_a\mu_a + x_b\mu_b)$ in both scenarios s_1 and s_2. In scenario s_1, the physician will leave the clinic at the end of rL. Thus, the physician's idle time is $(rL - x_a\mu_a - x_b\mu_b)$. However, in scenario s_2, the physician will leave the clinic either at the end of rL (if the last patient does not show up) or after the treatment of the last patient, which will be a number in the range of $[rL, T_2]$. Thus, the physician's idle time in Scenario s_2 is at least $(rL - x_a\mu_a - x_b\mu_b)$.

Therefore, in these two scenarios, the physician's idle time in scenario s_2 is no less than the physician's idle time in scenario s_1. For the same z vector, s_1 and s_2 share the same probability value $(1-p_a)^{x_a} p_a^{r_a - x_a} (1-p_b)^{x_b} p_b^{r_b - x_b}$. Since the physician's expected idle time will be the sum of 2^r scenario's weighted idle time, the physician's expected idle time in sequence π_2 is always no less than the physician's expected idle time in Sequence π_1. Note that for any schedule in set Λ, the expected physician overtime is zero. This completes the proof. ∎