2020

Impact of Performance-Based Budgeting on Quality Outcomes in U.S. Military Healthcare Facilities

Kimberly L. Decker Virginia Commonwealth University
Virginia Commonwealth University

Impact of Performance-Based Budgeting on Quality Outcomes in U.S. Military Healthcare Facilities

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

by

Kimberly L. Decker
MHA, Health Administration, Baylor University, 2012
B.A., Psychology, University of Virginia, 2004

Co-Advisor: Gloria J. Bazzoli, Ph.D.
Bon Secours Professor
Department of Health Administration

Co-Advisor: Christian Wernz, Ph.D.
Associate Professor
Department of Health Administration

Virginia Commonwealth University
Richmond, Virginia
Month, 2020

Acknowledgement

I am so very grateful for the many people in my life who supported me through this journey and, ultimately, made it possible. I could never have completed this process without the gift of your time, support, expertise, mentorship, and patience.

First, I want to thank my husband, Tim, for his faithful support of my career and education. Without the sacrifices he made in his own career, I would never have had the opportunity to fulfill this lifelong goal. He is the true definition of a partner, and I am so very honored to have him by my side. I would also like to acknowledge my children, Timmy and Lydia, who graciously coped with the sacrifices incident to having a mother in a full-time doctoral program. Completing this program required the full support of my family, and I cannot express how grateful I am for that gift.

I would also like to acknowledge my committee co-chairs, Dr. Gloria Bazzoli and Dr. Christian Wernz. I thank Dr. Gloria Bazzoli for her kind mentorship and detailed reviews of my dissertation. She dedicated countless hours of her limited time to my dissertation journey, and I cannot adequately express my gratitude for this gift. As she approaches retirement, she leaves behind a legacy of professionalism, kindness, and immense contribution to the field. I would also like to thank Dr. Christian Wernz for his dedication to the project and unique insight as an operations management researcher and industrial engineer. His mentorship and creativity helped me to develop my research idea and get the study started.

I would also like to acknowledge and thank my other two committee members, Dr. Stephen Schwab and Dr. Askar Chukmaitov. Dr. Schwab tutored me on Stata, taught me the nuances of working with DOD data, and helped me connect with all of the right people to apply for DOD data. His competence, mentorship, and generosity with his time enabled me to complete this dissertation without delay. Throughout this entire process, Steve has flawlessly fulfilled the dual roles of professional mentor and friend. I would also like to thank Dr. Chukmaitov for graciously stepping in on short notice and providing immediate, substantive recommendations at my proposal defense and beyond. Drawing from his valuable background as a physician and quality measurement researcher, he challenged me to think more deeply about the types of issues that may be most relevant to medical providers and the quality research domain. His perspective and recommendations helped me to significantly improve this research.

Finally, I would like to acknowledge Dr. Jan Clement and Dr. Dave Harless for helping me graduate. As the program director, Dr. Clement provided valuable mentorship throughout the entire doctoral process. Her calming influence and consummate professionalism helped me learn more about the art of scholarship and expectations of researchers in the field. She always set a positive example and helped me process interminable stacks of scholarship paperwork. I would be remiss if I did not also acknowledge Dr. Harless for his tireless efforts in teaching me econometrics. He never complained about me becoming a permanent presence in his office during office hours, and he continued his support to me throughout my dissertation data analysis. He never missed an opportunity to check in on me and ask about my family, which indicated to me that he genuinely cares for his students. His professionalism and positive example have given me several leadership lessons to take with me in the next steps of my career.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ACA	Affordable Care Act
ASD (HA)	Assistant Secretary of Defense for Health Affairs
CDC	Centers for Disease Control and Prevention
CLABSI	Central Line-Associated Bloodstream Infection
CMS	Centers for Medicare & Medicaid Services
DHA	Defense Health Agency
DHP	Defense Health Program
DID	Difference in Differences
DMIS ID	Defense Medical Information System Identification
DOD	U.S. Department of Defense
FY	Fiscal Year
GAO	U.S. Government Accountability Office
GME	Graduate Medical Education
GPRA	Government Performance and Results Act
HAC	Hospital-Acquired Condition
HEDIS	Healthcare Effectiveness Data and Information Set
HMO	Health Maintenance Organization
HVBP	Hospital Value-Based Purchasing
ICU	Intensive Care Units
IOM	Institute of Medicine
IRIS	Integrated Resource and Incentive System
LDL	Low Density Lipoproteins
MACRA	Medicare Access and Children's Health Insurance Program Reauthorization Act
MBO	Management by Objectives
MDR	Military Health System Data Repository
MEDCOM	U.S. Army Medical Command
MHS	Military Health System
MSM	Multi-Service Market
MTF	Military Treatment Facility
NCQA	National Committee for Quality Assurance
NDAA	National Defense Authorization Act
NHS	National Health Service
NHSN	National Healthcare Safety Network
OECD	Organization for Economic Cooperation and Development
OMB	Office of Management and Budget

PART	Program Assessment and Rating Tool
PBAM	Performance-Based Adjustment Model
PBB	Performance-Based Budgeting (also Performance Budgeting)
PPBS	Planning, Programming, Budgeting System
PPS	Prospective Payment System
PQRS	Physician Quality Reporting System
Premier HQID	Premier Hospital Quality Incentive Demonstration
PVBMP	Physician Value-Based Modifier Program
QOF	Quality and Outcomes Framework
QPP	Quality Payment Program
RDT	Resource Dependence Theory
RVU	Relative Value Unit
SIR	Standardized Infection Ratio
URI	Upper Respiratory Infection
VHA	Veterans Health Administration
ZBB	Zero-Based Budgeting

Abstract

IMPACT OF PERFORMANCE-BASED BUDGETING ON QUALITY OUTCOMES IN U.S. MILITARY HEALTHCARE FACILITIES

By Kimberly L. Decker, MHA

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

Virginia Commonwealth University, 2020

Director: Gloria J. Bazzoli, Ph.D.
Bon Secours Professor
Department of Health Administration

Co-Director: Christian Wernz, Ph.D.
Associate Professor
Department of Health Administration

Performance-based budgeting (PBB) is a variation of pay for performance that is applicable to government programs. It works by increasing or decreasing funding based on pre-established performance thresholds, which incentivizes organizations to improve performance. In late 2006, the U.S. Army implemented a PBB program in all of its healthcare facilities due to rising concerns over costs and quality in its facilities. The U.S. Army's PBB program tied hospital-level funding decisions to performance on key cost and quality-related metrics. This study examined the impact of this program and a subsequent PBB program on quality improvement in

U.S. Army healthcare facilities. Drawing from resource dependence theory, two hypotheses were developed, predicting that PBB would have a positive and sustained impact on quality performance in U.S. Army healthcare facilities. These hypotheses were tested using a retrospective difference-in-differences analysis of quality performance data in facilities exposed to PBB programs in comparison to Air Force and Navy facilities not exposed to PBB programs, both before and after program implementation. Data for this study were retrieved and merged from two data repositories operated by the Defense Health Agency in order to create a dataset encompassing a wide range of administrative, demographic, and performance information about 428 military healthcare facilities. The sample of 428 military healthcare facilities was divided into two groups based on exposure to the Army's PBB programs. Facility-level performance data on population health indicators and an inpatient clinical safety indicator were compared between the intervention group of 187 Army facilities participating in PBB programs and a comparison group of 241 Navy and Air Force facilities that did not participate in these programs. The study findings supported both hypotheses and suggest that the Army's PBB programs had a positive impact on quality performance. Facilities that participated in PBB programs increased performance after program implementation, relative to comparison facilities, for over half of the indicators under investigation. Furthermore, performance was evaluated for a 5-year period after program implementation for six quality measures. Performance in PBB facilities, relative to comparison facilities, was either sustained or continued to improve over the 5-year postperiod for five of the six performance indicators examined. Although this study has several limitations, the results are promising. The findings are relevant to clinicians and administrators in military and government-funded healthcare

organizations, as they offer evidence to support the future use of PBB as a mechanism for improving quality performance

Chapter 1: Introduction

This chapter is divided into six sections. The first two sections frame the study problem and its resultant research questions. The next two sections provide an overview of the theoretical framework and an analytic approach for the study. The fifth section discusses the significance of the study. The chapter concludes with an outline of the subsequent chapters in this dissertation.

Study Problem

The Institute of Medicine (IOM; 2001) report on the quality of the U.S. healthcare delivery system sparked almost two decades of research and experimentation to improve the quality of healthcare in the United States. Healthcare quality is a major concern for public and private payers alike, as medical errors and poor-quality care can increase the consumption of healthcare services and lead to excess costs (A. DeVries et al., 2012; Encinosa & Hellinger, 2008; Zhan & Miller, 2003). As a result, public and private payers have implemented a wide variety of financial incentive programs over the past 20 years to encourage healthcare organizations to deliver higher quality care.

Pay for performance is one approach for improving the quality of healthcare services. Pay for performance is an umbrella term used to describe the practice of providing financial remuneration linked to performance metrics (Glickman et al., 2007). The underlying logic is that when performance on an outcome metric is tied to financial incentives, organizations have a motive to pay more attention to that outcome and allocate internal resources to improve performance (T. L. Jones, 2018; Ocasio, 1997).

Pay for performance programs gained popularity among health insurers in the United States following demonstration programs sponsored by the Centers for Medicare & Medicaid Services (CMS) from 2003-2009 (James, 2012). In fiscal year (FY) 2013, CMS introduced a reimbursement model in which a small percentage of revenues were withheld from hospitals and had to be *earned back* through improvement on patient satisfaction and process of care measures (Petrullo, Lamar, Nwankwo-Otti, Alexander-Mills, & Viola, 2012). It has also been implemented internationally in countries such as England, where 70% of general practitioners are subject to pay for performance structures (Lester, Matharu, Mohammed, Lester, & Foskett-Tharby, 2013).

Though pay for performance has been widely implemented in U.S. healthcare systems, there has been little research consensus regarding its effectiveness. Pay for performance programs have been studied extensively in Veterans Health Administration (VHA) hospitals, public hospitals, and private hospitals, but these studies have yielded mixed results (Eijkenaar, Emmert, Scheppach, & Schöffski, 2013; Emmert, Eijkenaar, Kemter, Esslinger, & Schoffski, 2012; K. K. Kondo et al., 2018; R. Werner, Kolstad, Stuart, & Polsky, 2011). Specifically, pay for performance has only been shown to have a marginal positive impact on process of care measures and have little to no effect on outcome measures (Chee, Ryan, Wasfy, & Borden, 2016; Eijkenaar et al., 2013; Emmert et al., 2012; Houle, McAlister, Jackevicius, Chuck, & Tsuyuki, 2012; R. Werner et al., 2011).

Additionally, initial performance gains attributable to pay for performance incentives tend to diminish over time (Chee et al., 2016; K. K. Kondo et al., 2018; Petersen et al., 2013; R.

Werner et al., 2011). While some studies have documented pay for performance successes (Calikoglu, Murray, & Feeney, 2012; Huang et al., 2013; Lin, Yin, Huang, & Du, 2016; Yip et al., 2017), the overall evidence supporting its use is mixed and inconclusive (Damberg et al., 2014; Eijkenaar et al., 2013; Emmert et al., 2012; Van Herck, De Smedt, Annemans, Remmen, & Rosenthal, 2010).

A less common and less researched form of pay for performance is performance-based budgeting (PBB), which is specific to government programs (Kong, 2005). It is a policy mechanism that works by incorporating performance measures into the budgeting process in an effort to stimulate higher performance (Dunning, 2014; Kong, 2005). PBB enables key decision-makers to systematically account for the results achieved through public funding through the use of key performance indicators, a performance measurement system, and program evaluation (Dunning, 2014). The goal is to increase the efficiency and effectiveness of public expenditure by connecting organizational funding to specific results (Dunning, 2014).

In broad terms, PBB can be categorized as a specific form of pay for performance. Christianson, Leatherman, and Sutherland (2008) defined *pay for performance* as "any payment arrangement that specifically rewards quality" (p. 6S). Though PBB fits within this definition, a few key characteristics distinguish it from most traditional forms of pay for performance. First, PBB provides performance rewards almost exclusively at the organizational level. This contrasts with private sector pay for performance programs that have the potential to impact the individual pay of healthcare professionals. For example, the State of Georgia uses PBB to make decisions about funding particular department activities, which are then broken down into specific programs. Most of the performance measures used for these budget decisions relate to workload efficiency measures, such as number of clients served, number of cases completed, or proportion of various tasks accomplished (Lauth, 1985). These productivity data are submitted with budget requests and used for the purpose of allocating money in the budgeting process. Thus, the *reward* for strong performance is greater program funding, which occurs at the program or department level.

Another distinguishing feature of PBB is that the incentives are typically disbursed by a single government agency or organization. This enables managers to focus organizational attention on the metrics that are most important to the agency controlling the bulk of its financing. For the purposes of the current study, PBB is defined as a specific form of pay for performance that (a) applies to state or federally funded healthcare organizations and (b) incentivizes performance at the organization level through *budget mechanisms*.

In 2006, the U.S. Army implemented a PBB program in all of its healthcare facilities due to rising concerns over costs and quality in its facilities. The U.S. Army's PBB program tied hospital-level funding decisions to performance on key cost and quality-related metrics. Initial results from a pilot study demonstrated significant hospital-level improvements in productivity and quality performance measures (West & Cronk, 2011). However, PBB in U.S. Army healthcare facilities has not been systematically studied over time to determine if the initial effects have been sustained.

Research Questions

The purpose of this study is to investigate the effects of PBB on performance improvement in U.S. military healthcare facilities. The study's purpose is met through two research aims, both of which focus on the performance improvement in healthcare quality metrics. The first aim is to determine the impact of PBB on quality improvement in U.S. Army healthcare facilities. The second seeks to determine if quality improvements tied to PBB are sustained over time in those same facilities. Although the U.S. Army Medical Command's (MEDCOM's) PBB programs have targeted administrative, efficiency, and quality metrics, this study focuses solely on quality metrics due to the consistency with which these metrics have been measured and incentivized over the study period.

Research Question 1 draws upon observations from a wide body of research on pay for performance in order to examine whether similar patterns of results are observed for PBB programs. As mentioned, pay for performance programs have generated mixed results (Eijkenaar et al., 2013; Emmert et al., 2012; Kondo et al., 2018; Werner et al., 2011), varying from no effect to strongly beneficial (Damberg et al., 2014; Eijkenaar et al., 2013; Huang et al., 2013; Van Herck et al., 2010). Mixed results have largely been blamed on the heterogeneity of healthcare organizations and pay for performance incentive structures, (K. K. Kondo et al., 2018), but some of the variation can also be attributed to methodological differences across studies.

For example, the evidence for quality improvement in pay for performance programs tends to be weaker for studies with strong methodological designs in comparison to studies with weaker designs (Damberg et al., 2014; Eijkenaar et al., 2013). A few programs have demonstrated quality improvement in select measured areas (Lee et al., 2012; Weyer, Bobiak, & Stange, 2008), but some researchers have found that these improvements are not statistically better than results achieved through intensive quality improvement efforts without performance incentives (Glickman et al., 2007). Aside from methodological concerns, variations in findings of pay for performance studies are likely the result of design choices, context (Van Herck et al., 2010), patient characteristics, and provider characteristics (Chee et al., 2016). In light of these mixed findings in pay for performance studies, Research Question 1 examines whether PBB might demonstrate more promising effects in military hospitals.

> Research Question 1: What is the impact of performance-based budgeting on quality improvement in U.S. Army healthcare facilities?

Some researchers have questioned the long-term sustainability of performance improvement attributable to pay for performance incentives (Mendelson et al., 2017). Several studies have noted that initial performance improvements tend to attenuate or diminish over time (Bonfrer et al., 2014; Jha, Joynt, Orav, & Epstein, 2012; Van Herck et al., 2010; R. Werner et al., 2011). Specifically, the long-term effects of pay for performance can be impacted by factors such as an organization's baseline performance (K. K. Kondo et al., 2018; Lin et al., 2016; Markovitz & Ryan, 2017) and ceiling effects (Ryan, Blustein, Doran, Michelow, & Casalino, 2012). In light of these observations in pay for performance studies, Research Question 2 examines whether similar effects are observed in PBB programs.

> Research Question 2: Are quality improvements tied to performance-based budgeting sustained over time in U.S. Army hospitals?

Theoretical Framework

The current study employs the theoretical framework of Resource Dependence Theory (RDT; Pfeffer & Salancik, 1978) to examine the impact of PBB on quality metrics in U.S. Army medical facilities. The RDT framework has been used extensively in healthcare research related to financial incentives (Yeager, Menachemi, et al., 2014) and can therefore be useful in guiding this study on the effects of PBB incentives.

According to RDT, organizations depend on critical resources from the external environment to function and survive. Organizations must develop strategies to manage or reduce dependencies on those external resources critical to survival (Pfeffer & Salancik, 1978; Yeager, Menachemi, et al., 2014). RDT posits that external forces in the environment can influence the strategies that organizations adopt in order to manage access to critical resources (Pfeffer & Salancik, 1978).

Three environmental factors that affect organizations' access to critical resources are munificence, complexity, and dynamism. *Munificence* refers to the overall supply and accessibility of resources in the environment (Pfeffer & Salancik, 1978; Yeager, Menachemi, et al., 2014). Environmental *complexity* and *dynamism* are both related to the overall degree of uncertainty in the market (Yeager, Menachemi, et al., 2014). *Complexity* refers to factors that make strategic decision-making and actions more difficult (Pfeffer & Salancik, 1978). *Dynamism* refers to the level of change internal and external to the organization (Pfeffer & Salancik, 1978). The current study explores how these constructs can be applied within the context of the Military Health System (MHS) to explain and predict how PBB impacts quality performance in military healthcare facilities.

Analytic Approach

This study employs a quasi-experimental, post-hoc analysis of performance data in military healthcare facilities. Quality performance in facilities exposed to PBB programs are compared facilities not exposed to PBB programs, both before and after program implementation.

The data for this study are retrieved from two data repositories operated by the Defense Health Agency (DHA). The Military Health System Data Repository (MDR) is a centralized data repository that receives, archives, and validates DHA corporate healthcare data from military healthcare facilities worldwide (DHA, 2019f). The MDR uses standardized data processing methods to ensure that health data are collected and managed in a consistent manner across all U.S. Department of Defense (DOD) healthcare organizations (DHA, 2019f). The Carepoint Information Portal is also operated by the DHA and contains facility-level performance data for a wide range of quality measures (DHA, 2019b). It is the primary platform used by clinical leaders in MHS facilities to monitor and track performance trends for improvement initiatives.

The measures under analysis are commonly used in healthcare performance evaluation in the private sector. Measures include the Healthcare Effectiveness Data and Information Set (HEDIS; National Committee for Quality Assurance [NCQA], 2019) and the Centers for Disease Control and Prevention (CDC) National Healthcare Safety Network (NHSN) Program.

The study sample consists of 428 U.S. Military Treatment Facilities (MTFs) that are operated by the U.S. MHS. The MHS is a comprehensive, integrated health system that combines both direct and purchased care components to provide health services to eligible beneficiaries worldwide (*Defense Health Agency: 2017 Stakeholder Report*, 2017; DOD, 2014). All facilities in the sample are primarily operated by the medical departments of one of the three major service components (i.e., Army, Navy, or Air Force).

For Research Question 1, a difference-in-difference (DID) approach is used to study the effects of PBB on quality indicators in U.S. Army hospitals over a 16-year period from FY 2004 to FY 2019. During that period, U.S. Army hospitals implemented two successive PBB programs and incentivized various performance metrics at different points in time. Navy and Air Force healthcare facilities tracked performance on many of the same metrics but did not participate in PBB programs that tied their quality performance to hospital funding. Thus, performance trends in U.S. Army facilities (*n* =187) participating in PBB are compared to performance trends in a comparison sample of Navy (*n* = 153) and Air Force (n = 88) and combined-service facilities *not* participating in PBB programs. To assess performance trends on metrics prior to the implementation of the Army's PBB programs, performance data for each metric is examined for 2 years prior to incentivization. Additionally, data are examined for all metrics for 2 to 5 years after incentives are offered, depending on data availability. The overall effects of the PBB programs are assessed by comparing performance between the intervention and comparison groups in both the pre- and postincentive periods.

Research Question 2 addresses the sustainment of performance gains attributable to PBB programs. Performance trends are analyzed for two distinct postimplementation time periods: the first 2-1/2 years after incentivization, and the second 2-1/2 years postincentivization. This analysis is used to determine if performance trends are sustained for at least 5 years after the implementation of incentives and if there is an overall change in the pattern of performance over time (i.e., did performance decline, stay the same, or improve in the two postincentive time periods).

Significance

This study is significant due to its policy and research implications. From a policy perspective, improving the value of healthcare is a priority for the U.S. military. The scale of healthcare spending in this domain is enormous: the MHS is the second-largest healthcare system in the United States, operating at an annual cost of $50-$52 billion in service of 9.5

million military beneficiaries (DHA, 2019e; Kellermann, 2017; Mendez, 2018a). Over the past 20 years, healthcare spending in the U.S. military has risen sharply, consuming approximately 8-10% of the total annual defense budget (Mendez, 2018a), and raising concerns over long-term budget sustainability.

Healthcare quality is a particular concern within the MHS because active duty servicemembers must maintain peak health (i.e., medical readiness) in order to maintain military effectiveness and perform military duties in demanding environments (Hosek & Cecchine, 2001). This implies that the quality of healthcare that ill or injured soldiers receive potentially affects their availability to perform their military duties. It also generates an operational need for focused attention on population health and preventive care measures such as health screenings, tobacco cessation, weight management, vaccinations, and disease surveillance, since this can affect the overall readiness of the U.S. military force; therefore, the quality of military healthcare has been identified as a national defense priority (Pellerin, 2017). Despite this priority, performance-stimulating mechanisms such as PBB have not received much empirical attention. The current study addresses that gap by conducting a retrospective assessment of PBB in U.S. military healthcare organizations to determine its long-term effects.

The current study is also significant due to organizational changes that are occurring within the MHS. Over the next few years, the DHA will merge three separate healthcare systems (Army, Navy, and Air Force) into a single governance structure that will assume operational control over all U.S. military healthcare facilities (National Defense Authorization Act [NDAA], 2017). Since the DHA will have significant latitude to allocate resources among healthcare facilities across all three services, it is imperative for key decision-makers to have empirical evidence regarding the effects of PBB in military facilities. The mission of the DHA is guided by the MHS's Quadruple Aim of increased medical readiness (pertaining to the military force): better health, better care and lower cost (DHA, 2019c; Martin, Nelson, Lloyd, & Nolan, 2007). Therefore, it is important for leaders to know if PBB is an effective mechanism for quality improvement in U.S. Army healthcare facilities. If so, then leaders may consider expanding this policy to include all U.S. military facilities.

The current study is also important from a research perspective. In contrast to typical forms of pay for performance, there is little research on PBB and its effects in the healthcare field. The majority of the literature on PBB is derived from the political science domain and is not specific to healthcare organizations. There are very few experimental or quasi-experimental studies, so the effectiveness of PBB in comparison to other mechanisms has yet to be established. This study addresses that gap by providing a quasi-experimental approach to studying PBB in the healthcare domain.

The current study is also significant because it examines the impact of financial incentives in the underresearched context of military healthcare facilities. As Christianson et al. (2008) noted, the context in which pay for performance incentives are disbursed is a critical component to evaluating its effects. Contextual factors such as the size of the incentive (K. K. Kondo et al., 2018; Rosenthal & Dudley, 2007), the target of the incentive (individual, group, or organization; K. K. Kondo et al., 2018; Markovitz & Ryan, 2017; Petersen et al., 2013; Rosenthal & Dudley, 2011; Van Herck et al., 2010), and hospital characteristics (Damberg et al., 2014; Markovitz & Ryan, 2017; R. Werner et al., 2011) have all been shown to moderate the effects of pay for performance. It is important to expand the pay for performance research to include a wider variety of study settings to gain a more comprehensive understanding of the contextual factors that are associated with stronger (or weaker) pay for performance effects.

Additionally, there is considerable research on federally sponsored pay for performance programs from the perspective federal purchasers (such as Medicare and Medicaid); there are very few studies that address the effects of pay for performance from the perspective of federally funded healthcare providers. This study addresses those gaps by examining the effects of pay for performance in a military healthcare setting.

The current study also contributes to the pay for performance literature by examining the effects of performance-based incentives over an extended period of time. The initial results from a pilot study of the Army's PBB program demonstrated significant improvement in productivity and HEDIS performance measures (West, Cronk, Goodman, & Waymire, 2010). This study builds upon that evaluation to examine whether initial performance gains are sustained over a longer period of time. This is the first study of its kind to address the longitudinal impact of PBB in federally funded healthcare environments to determine if performance trends are similar to those observed in the private sector.

Summary of Remaining Chapters

This dissertation is organized into six chapters. This introductory chapter briefly outlined the thematic focus of the study, defined the research problem and research questions, introduced the conceptual approach and analytic approaches, and briefly explained the study's significance. Chapter 2 provides an overview of the relevant literature on pay for performance and PBB. It then transitions into an overview of the study setting, providing an explanation of the MHS and performance monitoring in military hospitals. Chapter 2 concludes with a description of the evolution of the Army's PBB programs. Chapter 3 discusses the conceptual model for this study, with an emphasis on RDT and its application in a military healthcare setting. The theoretical foundations of RDT are applied to derive two testable hypotheses.

Chapter 4 discusses the study's data, methodology, research design, and analytic approach. The results of the analysis are provided in Chapter 5, including the descriptive analyses, regression models, sensitivity analysis, and major study findings. Chapter 6 reflects on the results of the study, providing a discussion of its implications and limitations.

Chapter 2: Literature Review

This chapter is divided into four sections. The first section provides a broad overview of pay for performance, summarizing its history, uses and research findings. Section one also discusses the heterogeneity of research results in the pay for performance literature. The second section defines PBB and discusses its relevance to the pay for performance literature. The third section provides background information relevant to the context of this study in the MHS. The fourth section describes the evolution of the Army's PBB programs and the state of research into these programs. The chapter ends with a summary.

Overview of Pay for Performance

In broad terms, pay for performance is a financial arrangement in which payment is contingent upon performance. Within the healthcare domain, pay for performance is applied in an effort to improve clinical quality, increase efficiency, improve health outcomes, and enhance the patient experience (Damberg et al., 2014). The IOM (2007) described pay for performance as "the systematic and deliberate use of payment incentives that recognize and reward high levels of quality and quality improvement . . . and a powerful stimulus to drive institutional and provider behavior toward better quality" (p. 5). Thus, pay for performance is one part of an overall nationwide strategy to enhance the value of medical care in the United States.

Pay for performance initiatives have become increasingly popular in U.S. healthcare organizations over the last 20 years. This increase in popularity is partially attributable to two seminal reports published by the IOM: *To Err is Human: Building a Safer Health System* (2000) and *Crossing the Quality Chasm: A New Health System for the 21st Century* (2001). These two reports highlighted the fact that the U.S. healthcare system is not structured to provide safe, timely, and necessary care in support of the best possible health outcomes for patients. These reports asserted that under some circumstances, poorly delivered healthcare services can result in preventable patient harm or even death.

As a result, healthcare leaders have devoted the last 20 years to engaging a wide range of strategies aimed at reducing preventable harm and improving healthcare quality. Incentive programs have been directed toward a multitude of healthcare applications, including diabetes management (Gupta & Ayles, 2019; Huang et al., 2013), immunizations (Chaix-Couturier, Durand-Zaleski, Jolly, & Durieux, 2000; Kouides et al., 1998), hypertension (Petersen et al., 2013), ophthalmology (Herbst & Emmert, 2017), prenatal care (Rosenthal, Li, Robertson, & Milstein, 2009), reduction of hospital-acquired conditions (HACs; Bastian, Kang, Nembhard, Bloschichak, & Griffin, 2016; Calikoglu et al., 2012; Vokes, Bearman, & Bazzoli, 2018; Waters et al., 2017), cardiovascular conditions (Glickman et al., 2007), and chronic liver disease (Natarajan & Kanwal, 2015).

Pay for performance is just one of many strategies used to encourage healthcare providers to focus on quality improvement and reduction in preventable harm. Under pay for performance, payers adjust fee-for-service reimbursement rates on the basis of performance. Providers can either be rewarded (i.e., receive bonuses) or penalized (i.e., have payments reduced) for performance on pre-established targets for quality or efficiency (Damberg et al., 2014). For the purposes of the current study, this specific arrangement is distinguished from other types of performance-related risk-sharing agreements, such as capitated payments to accountable care organizations or bundled payments; however, it is important to note that there is a considerable amount of overlap in the research literature on all of these strategies (Damberg et al., 2014).

History and Evidence for Pay for Performance in the United States

The use of financial incentives to influence provider behavior and costs of care in the private sector dates back to the early 1990s (Damberg et al., 2014; Winslow, 1994) when private insurers sought alternatives to the traditional fee-for-service arrangements that rewarded providers based on the complexity and quantity of the services provided. Early programs in the United States started with small-scale efforts among a limited number of commercial insurers (Doran, Maurer, & Ryan, 2017; Rosenthal, Fernandopulle, Song, & Landon, 2004). These early programs were mostly implemented in managed care settings, in which providers were paid either through capitation or lump-sum arrangements for a given set of services (James, 2012). These payment arrangements motivated providers to control costs because they were sharing the financial risks of healthcare delivery with payers.

Pay for performance started to proliferate more widely in the United States in the early 2000s. Following the release of the two IOM (2000, 2001) reports highlighting safety errors and preventable harm in U.S. healthcare facilities, leaders and policymakers started seeking programs aimed at improving quality performance (James, 2012). Pay for performance emerged within this context as a method for incentivizing providers to improve quality and safety. The advent of the pay for performance was a major change for the U.S. healthcare industry (Dietrich, 2013; Rosenthal & Dudley, 2007). By 2006, over half of all health maintenance organizations (HMOs) in the United States included pay for performance in provider contracts (Rosenthal, Landon, Normand, Frank, & Epstein, 2006).

Pay for performance emerged in federal programs starting in 2003 with Medicare's flagship demonstration project, the Premier Hospital Quality Incentive Demonstration (Premier HQID). The initial phase of this voluntary demonstration project rewarded hospitals in the top two performance deciles for quality-related metrics pertaining to three medical conditions (acute myocardial infarction, congestive heart failure, and pneumonia) and two surgical procedures (coronary artery bypass grafting and total hip/knee replacement). Demonstration hospitals had the opportunity to earn a 1-2% bonus for high quality performance or (after 2006) be penalized for poor performance (Chee et al., 2016). Because the initial incentive structure did little to motivate low-performing hospitals to improve quality, the program was changed at the end of 2006 to reward quality improvement as well as performance relative to peers (Shih, Nicholas, Thumma, Birkmeyer, & Dimick, 2014).

Evidence for the effectiveness of the Premier HQID project is very limited. For example, Shih et al. (2014) found that the Premier HQID failed to improve surgical outcomes in participating hospitals. Similarly, numerous studies have found no greater reductions in mortality among demonstration hospitals in comparison to non-demonstration hospitals (Damberg et al., 2014; Glickman et al., 2007; Ryan, 2009). Lindenauer et al. (2007) found some limited evidence that pay for performance led to small improvements in *process of care* and *appropriate care* metrics in demonstration hospitals; however, they noted that much of this improvement could also be attributed to the effects of public reporting.

The Affordable Care Act (ACA) of 2010 served as a catalyst for further expansion of pay for performance in the United States. Numerous provisions in the ACA were designed to lower costs and improve the quality of healthcare delivery in the United States. Despite limited evidence to support their use, several programs were adopted to shift payments from volume-based reimbursement to value-based reimbursement predicated on quality performance (Doran et al., 2017). For example, Medicare's Hospital Value-Based Purchasing (HVBP) program incentivized quality performance by adjusting payments to hospitals based on how well they performed on either quality metrics relative to other hospitals or quality improvement over baseline (CMS, 2019b). Designed to be budget-neutral, the HVBP worked by withholding a portion of Medicare reimbursements to hospitals and then redistributing these funds among hospitals based on relative quality scores (Chee et al., 2016; Figueroa, Tsugawa, Jie Zheng, Orav, & Jha, 2016). The HVBP is modeled after the Premier HQID demonstration project, except participation in the HVBP was broader in scope and mandatory for all Medicare hospitals (Figueroa et al., 2016).

Early evaluations of the HVBP program did not demonstrate any more positive outcomes than the HQID demonstration project (Doran et al., 2017). For example, Ryan, Burgess, Pesko, Borden, and Dimick (2015) compared HVBP hospitals to critical access hospitals and hospitals in Maryland that were exempt from the program. They concluded that the HVBP hospitals demonstrated no greater improvement on process of care metrics or patient experience than exempt (non-HVBP) hospitals. Similar null effects were also observed for mortality associated with acute myocardial infarction, heart failure, and pneumonia (Figueroa et al., 2016).

The Hospital Readmission Reduction Program and the HAC Reduction Program are two additional programs sponsored by the CMS. These programs target quality improvement by reducing or eliminating payments to hospitals for adverse events or preventable hospital readmissions. For example, the HAC Reduction Program policy decrements reimbursements by 1% for the worst-performing quartile of hospitals nation-wide (CMS, 2019a; Vokes et al., 2018). The impact of this policy with respect to reducing HACs remains unclear (Vokes et al., 2018). Waters et al. (2017) found that the policy was associated with an 11% reduction in the rate of change in central line-associated bloodstream infections (CLABSI) and a 10% reduction in the rate of change in catheter-associated urinary tract infections. However, a similar study on infection rates following major urologic surgery did not find the rate of HACs to be affected by the policy (Rude et al., 2017).

The ACA also stimulated the adoption of pay for performance in outpatient care. For example, CMS initiated the Physician Quality Reporting System (PQRS) in 2011, rewarding physicians for reporting on a select set of quality metrics. This broadly scoped program included all physicians and allied health providers who billed Medicare for outpatient services (Natarajan & Kanwal, 2015). In 2015, CMS initiated the Physician Value-Based Modifier Program (PVBMP), which was closely linked to the PQRS. The PVBMP utilized the same basic reporting infrastructure as the PQRS, except that it applied penalties for not reporting quality data. It also expanded the scope of incentives to include performance-based penalties and bonuses (CMS, 2018; Chee et al., 2016).

In 2015, additional legislation set the conditions for the creation of the largest value-based purchasing program in the United States (Doran et al., 2017). The Medicare Access and Children's Health Insurance Program Reauthorization Act (MACRA) rescinded the sustainable growth formula previously used to establish physician payment rates. MACRA consolidated several quality programs into a single program known as the Quality Payment Program (QPP). The intent of this program was to change the physician payment scheme to better reward quality, value, and outcomes in healthcare services (CMS, 2015, 2020). The QPP rewarded quality using one of two methods: the Merit-Based Incentive Payment System or Advanced

Alternative Payment Models. Rigorous studies of performance change under QPP and MACRA are not yet available, so it is unclear what effect this legislation will have on healthcare quality in the United States.

International Pay for Performance Programs

In addition to the United States, pay for performance has become increasingly popular internationally, particularly among the developed countries that are members of the Organization for Economic Cooperation and Development (OECD; Feng et al., 2019; Milstein & Schreyoegg, 2016). Unlike the United States, most OECD countries have some form of universal healthcare coverage for its citizens (Dietrich, 2013). This factor is important because pay for performance programs can be more broadly implemented in single-payer systems in comparison to systems like the United States in which healthcare organizations cater to a diverse set of payers. No study has specifically compared the effects of pay for performance in countries with and without universal healthcare coverage, so it is unclear if this and other contextual factors may have the potential to alter the effects of pay for performance; however, due to differences in universal healthcare coverage, the effects of any international studies on pay for performance should be interpreted with caution when compared with studies in the United States. Despite this difference, international studies add to the overall depth of what is known to researchers about the effects of pay for performance in healthcare.

A significant portion of international research on pay for performance is derived from studies conducted in the United Kingdom and Taiwan, both of which provide universal healthcare coverage to their citizens. The Quality and Outcomes Framework (QOF) is one of the most widely researched programs, having been initiated in 2004 by England's National Health Service (NHS) in primary care settings. It was originally designed to boost recruitment in primary care by providing physicians with a substantial increase in pay (Roland & Campbell, 2014). In an effort to enhance performance on quality metrics, physician pay increases were made contingent upon performance in clinical indicators (i.e., condition-based care), organizational indicators (i.e., records, data, education and training, and practice management) and patient experience indicators (i.e., patient engagement surveys). Over the next 10 years, several changes were made to the performance indicators, including the addition of several more clinical indicators; the removal of organizational indicators; and the addition of public health indicators related to smoking, obesity, and sexual health (Roland & Campbell, 2014). The majority of studies on the QOF found positive performance effects, mostly on process of care metrics (Mendelson et al., 2017).

One of the most widely researched programs in Taiwan is the Diabetes Mellitus Pay for Performance Program initiated in 2001. Program enrollment was optional, and physicians were given the latitude to choose which patients to enroll (Chang, Lin, & Aron, 2012; Mendelson et al., 2017). This program started by incentivizing performance on process-of-care metrics but later transitioned to intermediate health outcome measures in 2006. Evidence from studies in this program is mixed, but most studies have documented positive effects on both process of care and intermediate health outcome measures (Mendelson et al., 2017).

International studies have generally reported positive effects of pay for performance in hospitals, though with decelerating trends in improvement over time (Kondo et al., 2015). International studies, particularly those in the United Kingdom and Taiwan, were more likely to show positive effects than studies within the United States (Mendelson et al., 2017); the reasons for these differences have not been established empirically.

Heterogeneity of Evidence on Pay for Performance

Overall, the evidence for pay for performance in the United States and abroad is contradictory and mixed. Some scholars have proposed that the lack of consensus may be the result of heterogeneity in patient and catchment area factors, organizational and structural capabilities, and program characteristics (Markovitz & Ryan, 2017). This heterogeneity is particularly visible in systematic review studies, which often incorporate findings from both single and multipayer systems, in various physician settings, and across multiple performance domains (Benabbas, Shan, Akindutire, Mehta, & Sinert, 2019). Due to these variations, it has been difficult for scholars to make broad generalizations about the overall success or failure of pay for performance (Allen, Mason, & Whittaker, 2014). Despite this, some context-specific trends have emerged in the research literature. Studies have shown that the effects of pay for performance may be sensitive to contextual influences such as care setting, patient factors, program design elements, and selection of quality metrics (Allen et al., 2014; Benabbas et al., 2019; Chen et al., 2017; Chien et al., 2012; Gupta & Ayles, 2019; Markovitz & Ryan, 2017). The following sections explore the impact of these factors in greater detail.

Care setting. Some of the heterogeneity observed in pay for performance studies may be attributable to differences in the care setting. There is considerable evidence that the effects of pay for performance are contingent upon the context in which incentives are introduced (Allen et al., 2014; Gupta & Ayles, 2019). Generally speaking, studies in ambulatory care settings have reported more positive results than studies in hospital settings, though the evidence in both settings has been mixed.

Some studies of hospital pay for performance have reported positive effects (Calikoglu et al., 2012), and others have reported negligible or no effects at all (Damberg et al., 2014;

Glickman et al., 2007; Shih et al., 2014). Hospital-based studies focused primarily on outcome metrics such as mortality, surgical outcomes, and HACs. Damberg et al. (2014) found that most mortality-focused studies offered no evidence that pay for performance reduced in-hospital mortality. Similarly, two studies on Medicare's Premier HQID project found no appreciable differences in quality improvement among pay for performance hospitals and comparison (non-pay for performance) hospitals.

Shih et al. (2014) studied mortality rates and surgical complications following cardiac and orthopedic procedures. They did not find an appreciable difference in mortality or complication rates following surgical procedures between demonstration and non-demonstration hospitals. Glickman et al. (2007) studied the effects of pay for performance on quality of care and outcomes for acute myocardial infarction. They compared 54 hospitals in the CMS Premier HQID project to 446 control hospitals that were not participating. They concluded that improvement in pay for performance hospitals was not incrementally better than improvement in comparison hospitals for quality of care or outcome measures.

In contrast to these studies, a study of two Maryland-based hospital pay for performance programs documented positive results. Calikoglu et al. (2012) found that all clinical process of care measures improved in the 4-year period following the implementation of value-based purchasing. They also found that HACs declined by more than 15% after implementation of a risk-adjusted pay for performance program.

Similar to hospital-based studies, evidence on pay for performance in ambulatory care settings is mixed (Allen et al., 2014). Findings varied based on the types of measures investigated and the quality of the studies. For example, there has been some limited evidence of success in diabetes care and disease management in primary care settings (Gupta & Ayles, 2019; Lin et al., 2016). Similarly, a systematic review on vaccinations found that eight out of nine studies in primary care settings noted a significant increase in vaccination rates following implementation of pay for performance incentives (Benabbas et al., 2019). In contrast to these findings, provider incentives were not found to have a large impact on other metrics, including well-baby care targeting low income patients (Felt-Lisk, Gimm, & Peterson, 2007).

According to two systematic reviews, variation in the results may be attributable to the quality of the study. For example, a systematic review by Scott et al. (2011) found modest, positive results in six out of seven studies, but the authors noted that many of these studies were low quality and subject to significant selection bias. Similarly, another systematic review that included a different set of ambulatory care studies concluded that higher quality studies tended to find either very small positive effects or no effects at all (Damberg et al., 2014).

Patient factors. Another source of heterogeneity in pay for performance results is the patient population. There is evidence that socioeconomic factors and medical risk in the patient population can impact facility performance. For example, in a study on Medicare's PVBMP, Chen et al. (2017) investigated facility-level performance on cost and quality indicators for physician practices serving medically and socially high-risk patients. Chen et al. found that physician practices serving more patients with high social risk factors had lower quality and lower costs, while practices serving a greater proportion of medically high-risk patients had lower quality and higher costs. Similarly, Chien et al. (2012) suggested that physician practices operating in economically depressed areas perform more poorly on pay for performance measures related to clinical quality, patient experience, and health technology usage.

Program design elements. Another factor that may explain the lack of consistency in pay for performance results is a variation in program design features. Pay for performance programs vary widely in the United States with incentive structures and program design features differing across payers and healthcare organizations. The most common form of pay for performance is a bonus or penalty adjustment that is attached to fee-for-service reimbursement rates. These rate adjustments are made on the basis of performance on various cost and quality indicators. For example, in FY 2013, CMS introduced a reimbursement model in which a small percentage of revenues were withheld from hospitals and had to be *earned back* through improvement on quality outcome measures and consistent patient satisfaction (Petrullo et al., 2012). Regardless of structure, the primary goal of pay for performance is consistent across programs: increase the value of healthcare purchasing by adding incentives for quality to complement payments for volume and complexity of services.

Though few studies have addressed these effects directly, there is evidence that program design features such as benchmarking, incentive target, incentive size, measure type, and provider engagement may play a role in obtaining the desired effects from pay for performance programs (Eijkenaar et al., 2013; Van Herck et al., 2010). The following sections explore these factors in greater detail.

Incentive size. Scholars have hypothesized that in order to impact quality performance, the size of the incentive must, at a minimum, be large enough for facilities to cover the costs associated with performance improvement efforts (Felt-Lisk et al., 2007; Werner & Dudley, 2012). Ideally, extrinsic rewards such as financial incentives should be large enough to motivate participation but not large enough to encourage undesirable behaviors such as exclusion of vulnerable or high-risk patients (Kondo et al., 2018).

A few studies have documented a relationship between incentive size and impact on performance (Damberg et al., 2014; Mullen, Frank, & Rosenthal, 2010; Werner et al., 2011).

Studies on England's QOF have also provided evidence regarding the effects of large incentives on quality performance. The QOF offered substantial quality bonuses to providers with physician practices having the opportunity to earn up to 1,000 quality points. Each quality point was worth £76- £130 (U.S. $133-$204), which provided a significant opportunity for physicians to boost their income. In 2009-2010, the average practice earned 937 points (range 878-972), resulting in approximately a 25% increase in individual physician income (Doran et al., 2006; Gillam, Niroshan Siriwardena, & Steel, 2012).

There is evidence that these large incentives worked. In a longitudinal study, Doran, Kontopantelis, and Valderas (2011) found significant increases in the rate of quality improvement for 22 of 23 quality indicators for the first year of implementation. Although the rate of improvement plateaued after the initial implementation period, the quality of care remained higher than pre-incentive levels for the remaining 2 years of the study. Improvements were also noted for staffing, documentation, and adoption of information technology (Roland & Campbell, 2014).

Some U.S.-based studies have also documented a relationship between incentive size and improvement. In a literature review, Damberg et al. (2014) found that larger incentives were positively associated with larger performance gains. Qualitative interview data from a VHA study suggested that the most effective way to impact clinically meaningful performance among physicians was to apply bigger, higher frequency incentives to performance measures that are within physician control (Kondo et al., 2018). Additionally, evidence from the Premier HQID demonstration project found that performance improvements were largest in facilities eligible for the largest bonuses (Werner et al., 2011).

Though economic theory predicts that the size of the reward impacts the degree to which organizations respond to incentives, the overall empirical evidence is inconclusive. Scholars have proposed a few explanations for the inconsistency in results (Markovitz & Ryan, 2017). In the United States, payer fragmentation may dilute financial incentives and reduce the impact of pay for performance programs (Van Herck et al., 2010). In multipayer systems, payer mix and number of patients for each payer impact the total financial incentive available to each healthcare provider. In some cases, the incentive may be very small. A study on Medicare's HBVP program found that performance incentives only changed payments by a fraction of 1% for two thirds of Medicare's participating hospitals, leading researchers to question whether such small incentives can significantly alter facility performance (Werner & Dudley, 2012). Another study on well-baby care in Medicaid highlighted the fact that incentives may not work if they insufficiently compensate providers for the effort required to access them (Felt-Lisk et al., 2007).

Though a few studies have suggested that larger incentives are more effective at changing behavior (Kondo et al., 2018; Van Herck et al., 2010; Werner et al., 2011), some healthcare leaders have criticized the use of large incentives on professional grounds. Their arguments are rooted in behavior theory, which proposes that extrinsic rewards such as monetary incentives may undermine the intrinsic motivation of healthcare providers (Deci, Koestner, & Ryan, 1999). This has the potential to degrade the altruism, compassion, and trust that are fundamental to health professions (Doran, 2014; Doran et al., 2017). Despite the perception from providers that pay for performance limits autonomy and professionalism, there is not yet conclusive evidence to support the assertion that pay for performance negatively impacts intrinsic motivation (Allen et al., 2014).

Selection of performance metrics. A key consideration for evaluating pay for performance programs is defining a common conception of clinical quality. The IOM (1990) defined *clinical quality* as "the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge" (p. 21). A multitude of different performance measures have emerged in the healthcare literature under this broad definition. This wide range of metrics presents a challenge for healthcare leaders attempting to design pay for performance programs that provide the maximum benefit to patients. As Jha (2017) noted, healthcare professionals often have difficulty achieving consensus regarding which metrics should be prioritized and by what degree. Inclusion of too many performance metrics in pay for performance programs may make it more difficult for organizations to prioritize their efforts on the measures that are most clinically beneficial to patients (Jha, 2017). This has the potential to dampen the effects of pay for performance programs and may be one factor explaining null effects in some pay for performance studies.

There are many domains of performance included in pay for performance programs. Early programs primarily focused on quality measures, but programs have recently expanded to include measures for both cost and quality. According to a RAND review of pay for performance (Damberg et al., 2014), categories of measurement typically include clinical processes, intermediate outcomes, patient safety measures, utilization, patient experience, outcomes, and structural elements. Though the healthcare quality literature contains many operationalizations of clinical quality, two of the most commonly used categories in the research literature are processes of care and outcomes of care (Hearld, Alexander, Fraser, & Jiang, 2008). *Process of care* metrics measure the degree to which healthcare providers follow established protocols in their treatment of patients. *Outcome metrics* indicate the impact the treatments and interventions have on the health status of patients (Agency for Healthcare Research and Quality [AHRQ], 2015). These two categories are consistent with the Donabedian classification paradigm for comparing healthcare quality on the basis of structure, process, and outcomes (AHRQ, 2015; Donabedian, 1980).

A great deal of literature has highlighted the difference in pay for performance effects between process metrics and outcome metrics (Damberg et al., 2014; Flodgren et al., 2011). For example, in a systematic review, Flodgren et al. (2011) found that financial incentives can be an effective means for changing physician behaviors in areas such as admissions, referrals, prescribing patterns, and processes of care; however, there is little evidence that these incentives improve patient outcomes (Allen et al., 2014; Flodgren et al., 2011). In a later systematic review, Damberg et al. (2014) compared process metrics and intermediate outcome measures and found mixed evidence for both. In general, higher quality studies have been less likely to identify performance improvement effects. Taken as a whole, most review studies have found evidence that process indicators yield higher performance gains than outcome measures in pay for performance programs (Damberg et al., 2014; Eijkenaar et al., 2013; Van Herck et al., 2010); however, effect sizes were small and results were mixed for both measure types.

Incentive thresholds. Performance thresholds for earning incentives can vary substantially among pay for performance programs. Some programs use absolute performance thresholds in which organizations must meet a fixed benchmark in order to obtain a bonus. Other programs rely on relative thresholds in which providers are assessed for performance relative to peers (Damberg et al., 2014). These two forms of benchmarking raise concerns that hospitals serving vulnerable and socioeconomically disadvantaged patients may be disproportionately penalized due to higher readmission rates, lower patient experience scores, and lower performance on process measures (Markovitz & Ryan, 2017).

Because physician organizations in socially disadvantaged areas score more poorly on pay for performance measures, some scholars have raised concerns that pay for performance may inadvertently increase resource gaps between organizations serving serve high- and low-income patients, thereby increasing health disparities (Chien et al., 2012). To address this concern, some programs have incorporated performance improvement incentives to help stimulate improvement efforts in low-performing facilities. Research evidence supporting the use of performance improvement incentives is limited. Within the Premier HQID program, performance improvement incentives were not shown to improve performance more rapidly among low-performing hospitals (Ryan, Blustein, & Casalino, 2012). A second study on Phase 2 of the Premier HQID program by Shih et al. (2014) yielded similar results. Expanding incentive opportunities for performance improvement was not shown to significantly impact surgical outcomes for cardiac and orthopedic procedures in demonstration hospitals. Despite limited evidence to support the use of performance improvement incentives, a study of a large insurer in Hawaii found that lower performing providers did respond well to small increases in the absolute performance thresholds (Bond, 2018). This study suggests that small increases in absolute performance thresholds may stimulate performance improvement in low-performing providers without raising costs.

Diminishing performance improvement effects. One consistent theme among pay for performance studies is that performance gains tend to diminish over time (Bonfrer et al., 2014; Kondo et al., 2018; Lin et al., 2016; Markovitz & Ryan, 2017; Mendelson et al., 2017; Ryan, Bluestein, & Casolino, 2012; Van Herck et al., 2010; R. Werner et al., 2011). One example is with the Quality Incentive Program in the United States. The early results of this program were positive: during the first year of incentives for end stage renal disease measures, 55-96% of facilities showed significant improvements on clinical process measures relative to baseline performance 2 years earlier (VanLare & Conway, 2012). However, many of the positive results captured in the initial program evaluations did not hold up in longer term studies. A 5-year study of the Premier HQID program found that more demonstration hospitals achieved high performance scores in the first few years after the implementation of pay for performance than in comparison hospitals not participating in pay for performance. However, the effect diminished over time, and differences disappeared after 5 years (Werner et al., 2011). Similarly, another 6-year study on long-term effects of pay for performance in inpatient settings found no evidence for a long-term effect on 30-day mortality (Jha et al., 2012). Comparable results were also observed for long-term studies of screening and preventive care in primary care settings (Kondo et al., 2018).

Despite a general lack of research consensus regarding the effects of pay for performance on outcome measures, one noteworthy study did find positive effects. A study on a hospital-based program in Pennsylvania found that participation in pay for performance was associated with a 27% reduction in the rate of CLABSIs (Bastian et al., 2016). Researchers also noted significant effects based on the length of time the hospitals participated in the program. For example, long-participating hospitals had an average of 3.13 fewer CLABSIs per year in comparison to those participating for fewer than 4 years (Bastian et al., 2016). This study is noteworthy because it suggests that length of program participation may be an important consideration for researchers evaluating the effects of pay for performance on outcome measures. Evaluation periods that are too short my run a risk of failing to fully capture significant effects.

A long-term study within the VHA provided a notable exception to the pattern of diminishing performance effects observed in most pay for performance studies. This study found significant and sustained improvements on process measures for six out of seven indicators within VHA facilities participating in pay for performance (Benzer et al., 2014; Kondo et al., 2018). This unique finding lends itself to additional research, particularly regarding the long-term effects of pay for performance in federal healthcare facilities. It raises questions about the potential influence of contextual factors such as hospital type (e.g., federal versus privately operated hospitals) on the effects of pay for performance.

PBB: A Special Case

The management concepts of pay for performance can also be applied in government programs. This special form of pay for performance is known as *performance-based budgeting* (PBB). Though there are many forms of PBB discussed in the following sections, the basic mechanism for all forms of PBB is to provide a financial motivation for public programs and government-funded organizations to improve and maintain performance. This is achieved by establishing performance criteria and using data on these criteria to assess performance and allocate budget resources. Robinson and Brumby (2005) defined PBB as “procedures or mechanisms intended to strengthen the links between funds provided to public sector entities and their outcomes and/or outputs through the use of formal performance information in resource allocation decision-making” (p. 15). PBB is a policy mechanism that enables decision-makers to systematically account for the results achieved with public funds (Dunning, 2014). The goal of PBB is to increase the efficiency and effectiveness of public expenditure by connecting organizational funding to specific results (Dunning, 2014). Because the theoretical basis of PBB is closely linked to pay for performance, the following sections discuss its definitions, history, and similarities to pay for performance. Key distinctions from pay for performance are also discussed, as well as its application in healthcare settings.

Definitions

The academic literature on PBB introduces a range of definitions and applications. PBB has been described as a *budgeting system* in which input costs are related to performance (DeVries & Nemec, 2019, p. 5) and funding allocation is related to measurable results (OECD, 2003). It has also been described as a *process* involving the use of performance information for budgetary purposes (Mauro, Cinquini, & Grossi, 2017; OECD, 2007). Jordan and Hackbart (1999) distinguished between performance budgeting and performance funding, defining the former as “preparing a budget document that includes performance information” and the latter as “the allocation or distribution of a percentage of appropriated funds contingent upon the assessment of the performance measures identified in the budget” (p. 69). Using these distinctions, performance funding is a more regimented form of PBB because it allocates money on the basis of performance thresholds rather than just using performance information while making budget decisions.

The academic literature delineates two separate dimensions for measuring performance in PBB programs (DeVries & Nemec, 2019; Robinson & Brumby, 2005). The first dimension, *outputs*, focuses on the organizational activities involved with delivering a product or service (Kong, 2005, p. 97). Output-based performance measurement is intended to achieve allocative efficiency. In the economic literature, allocative efficiency refers to an optimal distribution of goods and services in which the output of production is as close as possible to the marginal cost

(Agarwal, 2019). In the PBB literature, allocative efficiency refers to funding programs that provide the maximum benefit for public expenditure (Robinson & Brumby, 2005). This is accomplished by linking funding to quantifiable activity measures such as relative value units (RVUs), the number of patients seen, or the quantity of healthcare services delivered. Performance is defined in terms of outputs as a function of budget inputs.

The second dimension of PBB gauges performance through the use of *outcome* measures. Outcome-based measures address the question of whether or not a program achieves its intended goals (Kong, 2005, p. 97). Outcome measures are used to assess the quality or effectiveness of programs and services and are therefore the most useful to policymakers in budget decisions (Kong, 2005). In contrast to output measures that focus on what is bought through resource expenditures, outcome measures focus on what is actually achieved through resource expenditure (M. S. DeVries & Nemec, 2019).

History

PBB is purported to have originated in the U.S. Congress in the late 1940s, though some scholars have argued that it has existed in concept since the rule of Chinese Emperor K'ang-shi in the 1700s (DeVries & Nemec, 2019). The U.S. experience with PBB started under President Truman in the late 1940s. The Hoover Commission, a body charged with recommending administrative changes to the executive branch of government, proposed changes to the federal budgeting processes. The proposed changes shifted budgeting from its traditional line-item expenditure format to a new *performance budget* that emphasized functions, costs, activities, and accomplishments of federal expenditures (U.S. Government Accountability Office [GAO], 1997). In 1949, the National Security Act Amendment prompted the implementation of performance budgeting in the U.S. Military (Jordan & Hackbart, 1999). Shortly after, Congress passed the Budget and Accounting Procedures Act of 1950, which modernized federal accounting procedures by requiring the president to submit "functions and activities" of the government with the budget request; this new format shifted focus from inputs (expenditures) to outputs of the federal government, such as programs, weapons. and training (McGill, 2001, pp. 377-378). The Hoover Commission reforms were critical to the development of PBB in the United States because they established the philosophical underpinnings of a new performance-oriented approach to budgeting that would evolve over the next 70 years (Jordan & Hackbart, 1999, p. 68; GAO, 1997).

The Hoover Commission reforms were followed by a series of legislative actions that incrementally transformed budgeting and managerial practices in public sector agencies in the United States. The first evolution occurred in 1965 with the initiation of President Johnson's Planning, Programming, Budgeting System (PPBS). The PPBS established a framework for executive branch decision-making that incorporated an analysis of long-term policy objectives in the budget formulation process (GAO, 1997). Similar to the Hoover Commission's budget legacy, under the PPBS, *performance* was mostly measured in terms of government outputs (GAO, 1997). A critical aspect of PPBS was its use of sophisticated analytical tools to link government outputs to long-term policy objectives. This complexity of analysis, combined with its inability to account nonquantifiable political factors, ultimately led to the replacement of the PPBS form of budgeting (McGill, 2001, p. 378).

The next evolution in U.S. PBB occurred with the introduction of President Nixon's Management by Objectives (MBO), which was intended to hold managers accountable for the outputs of their organizations by linking the stated objectives of agencies to their budget requests (McGill, 2001). Like previous programs, performance under MBO was mostly measured in terms of outputs and efficiency measures; however, it also included some early efforts at assessing program results and outcomes (McGill, 2001). Both PPBS and MBO were important points of evolution in PBB because they championed the use of sophisticated performance measurements and analysis in spending decisions (GAO, 1997).

Zero-Based Budgeting (ZBB) was introduced into the executive budgeting process in 1977 by President Carter. ZBB required government agencies to rank order priorities based on alternative spending levels without the use of a budget base from previous fiscal periods (GAO, 1997). This structure required agencies to re-analyze and re-prioritize expenditures with each fiscal cycle. ZBB was an important step in the PBB evolution in the United States because it was the first to pursue a specific connection between budget resources and program results (GAO, 1997), effectively forcing agencies to make hard choices between priorities at alternative spending levels (McGill, 2001).

The most recent evolution in PBB is the Government Performance and Results Act (GPRA/Results Act) of 1993, which was designed to strengthen federal decision-making and accountability while helping Congress gain better visibility of program results in relation to government expenditures (GAO, 1999). The Results Act shifted the focus of accountability from processes to results, requiring federal agencies to submit 5-year strategic plans outlining the outcomes they hoped to achieve with government resources ("Linking Program Funding to Performance Results," 2002). It also required agencies to submit annual performance plans detailing performance goals and plans for achieving them. The Results Act was a significant advancement to progress in PBB in the United States because it explicitly required each agency to incorporate performance goals and program activities to their budget requests (GAO, 1999).

It was also significant because it relied on the use of objective information to determine the relative effectiveness and efficiency of government programs.

By 2002, several analyses indicated that federal policymakers had largely failed to adequately establish a link between performance results and their funding decisions ("Linking Program Funding to Performance Results," 2002; GAO, 1999) as intended by the Results Act. In response to this issue, the Office of Management and Budget (OMB) developed the Program Assessment and Rating Tool (PART) in 2002. The PART was designed to help agencies assess programs more objectively. It also aided agencies in developing reasonable goals and measures, feasible strategies to achieve their goals, and credible performance information ("Linking Program Funding to Performance Results," 2002, p. 7). The PART assessment included four basic questions, including the following:

1. Does the program have a clear definition of success, and is it designed to achieve it?
2. Are the program goals sufficiently outcome-oriented and aggressive?
3. Is the program well managed?
4. Does the program achieve its goals? ("Performance-Based Budgeting," 2005, p. 5)

Programs were awarded points based on the answers to each of these questions. After adding all of the points, each program was rated according to five categories: effective, moderately effective, adequate, ineffective, and results not demonstrated. In addition to making the ratings public, Congress started to use ratings in budget deliberations. As of 2004, 29 programs were terminated or reduced due to receiving a rating of Results Not Demonstrated ("Performance-Based Budgeting," 2005, p.6).

The PART was used by Congress until 2008 when it was replaced by performance initiatives crafted by the Obama Administration. In 2010, the Obama Administration enacted legislation that updated the GPRA of 1993. The GPRA Modernization Act of 2010 (Pub. L. 111-352) placed emphasis on incorporating performance into federal program management. This legislation not only encouraged the use of performance information in budgeting, it also encouraged federal managers to incorporate objective performance information into the overall management of federal programs.

In aggregate, the evolution of PBB in the United States since the 1940s demonstrates a persistent effort to broaden the use of performance information in the budgeting and management of federal programs. While there is a consensus among scholars that these programs have often failed to achieve the goals of their design (McGill, 2001; GAO, 1997, 2018b), each transformation moved the practice of PBB forward in a visible way (Kong, 2005). PBB started in the late 1940s with the Hoover Commission's recommendation to emphasize the outputs of government funds rather than the expenditures alone. By the 1990s, PBB evolved to emphasize the goals of measuring performance in terms of outcomes rather than outputs. By 2003, the GAO (2018b) found that government leaders had greater access to performance information than they had in the late 1990s. Despite this forward progress, several recent government reports still call for PBB programs to more concretely link performance to budget decisions (OMB, 2011; "Linking Program Funding to Performance Results," 2002; "Performance-Based Budgeting," 2005; GAO, 2018b). Thus, PBB programs continue to undergo evaluation and change. Though PBB research continues to evolve, the following section provides an overview of what is currently known about the effectiveness of PBB in government programs.

Key Findings About PBB

Because budgeting is a fundamentally political action, the academic literature on PBB is mostly situated in the political science domain. Budgeting is a political action because it forces policymakers to exercise political choice in allocating resources among a range of alternatives (GAO, 1999). PBB infuses performance data into this inherently political process. Most of the research literature on PBB is supported through case studies, government reports, literature reviews, and international comparisons. Unlike the management and healthcare literature, there are very few experimental or quasi-experimental designs in the PBB literature.

Though there have been a few documented successes, many PBB program attempts have failed (Jordan & Hackbart, 1999; McGill, 2001; McNab & Melese, 2003; GAO, 1997). According to a review by McNab and Melese (2003), these failures are attributable to "administrative complexities, lack of investment in managerial, accounting and information systems, and the lack of institutional incentives to promote gains in economic efficiency" (p. 73).

One consistent problem is the failure to use available performance information in the budgeting process. Numerous reviews of PBB programs in the United States have indicated that even when performance information is available, most programs fail to explicitly incorporate that information into budget decisions (Coplin, Merget, & Bourdeaux, 2002; Kong, 2005; "Linking Program Funding to Performance Results," 2002; GAO, 2018b). A GAO (2018b) survey found that while state government officials reported greater availability of performance information since 1997, there had been little to no change in the use of that information for various management activities (p. 6). Similarly, another GAO survey found that government-wide use of performance information did not improve after the enactment of the

2010 GPRMA Modernization Act. Though managers had access to more performance information, they did not use that performance information in program management any more often in 2017 than in 2013 (GAO, 2018b). This trend was also noted by Representative Stephen Horn, who opened a 2002 Congressional hearing on performance budgeting by stating that "policymakers have failed to establish a connection between performance results and their funding decisions . . . and the effectiveness of funding decisions [is] largely untested" ("Linking Program Funding to Performance Results," 2002).

Though performance information is more widely available to policymakers now than in the 1980s and 1990s, many leaders have still cited difficulty obtaining and communicating credible performance information. According to a GAO survey conducted in 2002, only five of the government's 24 largest agencies could attest to the completeness and reliability of the information used in budget decisions ("Linking Program Funding to Performance Results," 2002). One factor is that highly useful information is often costly or difficult to obtain (Schick, 2003, p.74). The costs associated with implementing a sophisticated performance monitoring system can be high in terms of manpower, additional documentation requirements, and information technology for data mining. This contributes to the pervasive view that the administrative costs outweigh its benefits (Kong, 2005, p. 93). Despite the high cost, when resources are allocated to data collection, it impacts the degree to which performance information is used in budgeting. For example, Jordan and Hackbart (1999) found a positive association with the number of analysts in a budget office and the state's use of performance information in funding processes.

Another complication for PBB is that institutional, functional, policy, and political constraints are fundamental to the American decision-making system (Radin, 2000, p. 133).

This means that performance is not the only information considered in budgeting decisions, even when sophisticated PBB programs exist. Dongsung Kong (2005) noted,

> Public budgeting at all levels of government is intrinsically political. Performance measures, or any rational ideas, will not supersede political priorities in any near future. PBB can be appropriate and applicable to managerial and less political decisions, but not highly political decisions. (p. 103)

This means that even when highly developed analytical tools are used to review performance data during budget deliberations, political forces can eclipse objective analyses.

For example, in the early 2000s, Congress used the PART to affect change in the budget process for the Community Development Block Grant Program. Despite its intended analytical approach, Representative Michael Conaway testified that Congress "did not get it done in the budgetary process. The political backlash, the whole ownership of those particular programs overran the analysis piece of what was going on" ("Performance-Based Budgeting," 2005). As some other members of Congress have noted, performance information in PBB programs do not fully account for all of the factors that must be considered in budget deliberations. For example, according to testimony by Clay Johnson of the OMB, high priority activities (such as certain defense programs) must sometimes be funded regardless of performance ("Performance-Based Budgeting," 2005).

Deciding upon an appropriate measurement system can also be politically problematic. According to congressional testimony from James Nussle, Chairman of the House Budget Committee, one of the key challenges with implementing the PART was determining how performance would be measured and who would decide which performance measurements would be used ("Performance-Based Budgeting," 2005). When PBB applies to politically sensitive programs, there may also be political interest in defining a particular measurement approach.

Interpreting performance information is another factor that complicates the implementation of PBB programs. Though outcomes-based performance measures are widely accepted as the most desirable for PBB programs (Schick, 2003, p. 74), these data can often be difficult to interpret for budgeting purposes. For example, it is sometimes complex to disentangle the effects of multiple programs on a particular outcome (Kong, 2005). When multiple programs contribute to a particular outcome, it can be difficult to make an allocation decision. Williams and Melhuish (1999) noted, "with no means to measure an individual's contribution, a system of 'sticks and carrots' is arbitrary and of little impact" (p. 25).

For all of the reasons mentioned, PBB programs have evolved over time in the United States. PBB programs have been redesigned through multiple legislative iterations, suggesting that certain aspects of these programs may not have worked smoothly in the intervening years (Schick, 2003). Like pay for performance, the effects of PBB programs are often moderated by factors outside of the program itself. PBB programs are difficult to implement if they are overly complex, include performance measures that are outside of managerial control, or fail to explicitly link performance to financial incentives. In contrast to pay for performance, PBB programs have the additional complication of being influenced by political forces that threaten to undermine their analytic approach. The following section provides a more detailed comparison of PBB and pay for performance within the domain of healthcare.

Comparison of PBB and Pay for Performance in Healthcare

PBB programs in federal healthcare facilities are very similar to traditional forms of pay for performance. PBB incorporates many of the same basic mechanisms as pay for performance but applies them to government-sponsored programs rather than private sector organizations. Christianson et al. (2008) defined *pay for performance* as “any payment arrangement that specifically rewards quality” (p. 6S). Under this broad definition, PBB can be categorized as a pay for performance program, though three key distinctions exist. First, government organizations are not subject to the same market forces as private sector organizations (Williams & Melhuish, 1999; Zemrani, 2019). In the private sector, market forces are the catalyst for performance improvement. Businesses must address quality and satisfaction concerns of patients and customers to remain viable (Zemrani, 2019, p. 36). Similarly, hospitals and healthcare organizations compete with each other in terms of the quality of their healthcare services and patient amenities in order to attract business and remain financially viable. In the public sector, these market forces are less pronounced or absent. The ultimate measure of success for government programs is the service that is delivered through public expenditure not its profit margin (Williams & Melhuish, 1999, p. 23). When publicly funded organizations have stable funding that is not tied to performance, they have less incentive to improve performance and address satisfaction concerns. By linking performance to the level of resourcing, PBB provides a potential mechanism for policymakers to hold organizations accountable for service quality and costs.

A second distinction between PBB and pay for performance is pay for performance programs have the potential to impact individual pay, whereas most PBB programs usually do not. Many U.S. pay for performance programs employ quality-based bonus payments that supplement capitation rates or fee-for-service reimbursements (Christianson et al., 2008). Payment arrangements and performance thresholds are usually stipulated in contracts between purchasers and healthcare organizations, but organizations can also incentivize individual physicians by offering performance bonuses (Rosenthal & Dudley, 2007). U.S. physicians and physician groups can earn up to 10% of their annual income in pay for performance bonuses, although the amount is typically less (Christianson et al., 2008; Rosenthal & Dudley, 2007; Sherry, 2016). In VHA facilities, pay for performance bonuses are disbursed at the organization level through a central VA office, and facility-level senior managers then have the ability to distribute bonuses to individual providers and employees (Benzer et al., 2014). In contrast to pay for performance, PBB incentives usually remain at the organizational level. Incentives are disbursed through budget mechanisms (increased or decreased program funding) and have less potential to directly impact individual pay.

A third difference for PBB in the healthcare domain is that incentives are typically controlled by a single government agency rather than a collection of individual payers. Since federal hospitals are almost exclusively financed through public expenditure, organizations are usually dependent upon a single agency for the majority of their operational resources. Obtaining critical resources through a single agency reduces the complexity associated with tailoring performance improvement initiatives to suit the requirements of multiple payers. With fewer competing demands, organizational decision-makers may be more able to focus organizational attention on the performance metrics that are most important to the agency controlling the bulk of its funding and financial resources (Ocasio, 1997).

In contrast to publicly funded programs, commercial healthcare organizations must manage performance targets from a conglomerate of healthcare purchasers. For example, an analysis of 48 state and regional health measure sets identified 509 distinct metrics with only 20% overlap of metrics between programs (Bazinsky & Bailit, 2013). If performance measures do not overlap across multiple insurers, organizations must expend additional resources on measuring and reporting across many different metrics. As Cassel et al. (2012) noted, this can lead to measurement fatigue without a corresponding change in results.

As mentioned, having too many metrics makes it difficult for managers to focus organizational attention on the metrics that matter the most to patient outcomes (Cassel et al., 2012; Ocasio, 1997). As a result, it is possible for organizational efforts to focus on minimally beneficial metrics rather than measures that may have a greater impact on patient outcomes. This is a factor that may account for the relative lack of success in many U.S.-based pay for performance programs in comparison to programs in single-payer countries such as the United Kingdom (Mendelson et al., 2017). Also, incentives from single payers are potentially larger because they are applicable to the entire population of patients rather than being segmented among patients based on patient insurance coverage.

Due to the aforementioned distinctions in market forces, incentive source (government funding versus insurance payments), and incentive impact (organizations versus individuals), PBB should be considered a *special case* of healthcare pay for performance that is specific to government-financed healthcare organizations. For the purposes of this dissertation, PBB is defined as a specific form of pay for performance that (a) applies to state or federally funded healthcare organizations and (b) incentivizes performance at the organization level through *budget mechanisms*.

Overview of the MHS and Links to PBB

One U.S. government organization to use PBB is MEDCOM, which is one component of the U.S. MHS. MEDCOM experimentation with PBB started around 2004 with a small pilot program conducted in a handful of U.S. Army medical facilities in the southeast region of the United States. As of 2019, the DHA is in the process of expanding PBB to all other service components within the MHS. The following two sections discuss that program and the broader context of the MHS, which set the conditions for its inception.

The MHS is one of the largest health systems in the world, delivering medical and dental care to military beneficiaries in almost 700 hospitals and clinics worldwide (Smith, Bono, & Slinger, 2017). The MHS is a comprehensive, integrated health system that combines both direct and purchased care components to provide health services to eligible military beneficiaries (*Defense Health Agency: 2017 Stakeholder Report*, 2017; DOD, 2014). As of 2018, the MHS employs a total of over 240,000 medical staff, split between active duty (48%) and reserve (30%) personnel, federal civilians (18%), and contractors (4%) (GAO, 2018a). The scope and scale of MHS healthcare services are substantial. The MHS provides services to 9.51 million eligible patients, including 1.4 million active duty servicemembers and 8.1 million reservists, retirees, and family members and survivors (DHA, 2019e; DOD, 2014; Weil, 2019). In 2018, MHS hospitals documented over 204,000 inpatient admissions, 39 million outpatient encounters and 1.17 million emergency department visits (Adirim, 2019).

Patient care is provided in the MHS through two integrated components (Childress, 2013, p. 7). The direct care component is comprised of health facilities directly operated by agencies within the DOD. The direct care system in the MHS is analogous to a staff-model HMO that directly employs salaried providers (Bond & Schwab, 2019, p. 1328). The purchased care component is comprised of a network of contracted civilian healthcare providers who offer a critical supplement to the healthcare services delivered within the MHS direct care system. These supplemental contract services allow the MHS to provide access to beneficiaries when MTFs are over capacity or beneficiaries live outside of the catchment area of military healthcare facilities. They also provide a referral option when the scope of services required exceed the services available through local military hospitals (Tanielian & Farmer, 2019). Sixty percent of DOD health services are delivered through these contracted providers (Adirim, 2019, p. 1269).

TRICARE is the health plan that integrates the direct care and purchased care components. Chapter 55 of Title 10 U.S. Code authorizes the entitlement of TRICARE health benefits as a component of the compensation package afforded to active duty military personnel, their families, and retirees. Eligible beneficiaries can participate in three main benefit plans. TRICARE Prime is a HMO option with very little cost sharing. TRICARE Select operates like a preferred provider organization, in which beneficiaries are offered more choices for civilian healthcare providers but often pay higher cost shares. TRICARE for Life is a Medicare wrap-around plan that is offered to Medicare-eligible retirees. The TRICARE program also includes pharmacy benefits and optional vision and dental plans (Mendez, 2018a, p. 10). TRICARE benefits and available plans vary slightly with respect to priority of access to military facilities and degree of cost sharing. This is largely based on the beneficiary category (i.e., active duty, dependent, retiree, retiree dependent, reserve, national guard). For example, active duty beneficiaries have no out-of-pocket costs, while family members of active duty incur cost-sharing expenses only when using out-of-network care without a referral (Tanielian & Farmer, 2019).

10 U.S. C. §1074 defines priorities for access to care in MTFs. Active duty servicemembers are the only beneficiary group entitled to care in MTFs (10 U.S. C. §1074). Space available priority for access to MTF care is then provided (in order) for active duty family members enrolled in TRICARE Prime, retirees, and their family members enrolled in TRICARE prime; active duty family members not enrolled in TRICARE Prime; and all other eligible persons (32 C.F.R. §199.17(d) - TRICARE program, n.d.; DOD, 2011).

Coverage and benefits have expanded in the TRICARE program over the last 20 years due to their necessity in attracting and maintaining an all-volunteer military force during a period of extended military conflict. Cost-sharing has remained very low relative to civilian insurance programs (Dolfini-Reed & Jebo, 2000; Military Compensation and Retirement Modernization Commission, 2015; Tanielian & Farmer, 2019). Because these benefits constitute a substantial investment of federal funds, members of Congress work closely with key leaders in the DOD and MHS to ensure that the medical needs of military members, retirees and military families are met through the MHS and TRICARE infrastructure. This is evidenced by the fact that Congress has updated military healthcare benefits in every FY since 1976 (Dolfini-Reed & Jebo, 2000). Though Congress maintains responsibility for defining eligibility criteria and scope of benefits under the TRICARE program, the MHS establishes the contracts, policy, budget, oversight, and civilian provider networks necessary to deliver this care (Hutter et al., 2019; Tanielian & Farmer, 2019).

Mission of the MHS

The mission of the MHS is to

> ensure America's 1.4 million active duty and 331,000 reserve-component personnel are healthy so they can complete their national security missions; to ensure that all active and reserve medical personnel in uniform are trained and ready to provide medical care in support of operational forces around the world; and to provide a medical benefit commensurate with the service and sacrifice of more than 9.4 million active duty personnel, military retirees and their families. (DHA, 2019a)

Commensurate with this mission, the MHS has two critical objectives—often referred to as the *readiness mission* and the *benefits mission* (Tanielian & Farmer, 2019). The primary objective of the MHS is the readiness mission. The readiness mission involves maintaining the health of military personnel so they can carry out their military duties and deliver healthcare services during military operations (Mendez, 2018a; DOD, 2014). In support of this readiness task, the MHS provides preventive and restorative health services to military personnel so that they can be medically prepared to deploy at a moment's notice. U.S. military forces are deployed worldwide in support of diverse missions such as combat operations, responses to natural disasters, humanitarian intervention, training, deterrence, and diplomacy (Hutter et al., 2019). These missions require servicemembers to be at peak health to withstand the physical and mental strains associated with working in austere and demanding operational environments.

As part of its readiness mission, the MHS is also charged with maintaining a *ready medical force* of military healthcare providers who are trained and equipped to provide medical care in a complex operational environment (Hutter et al., 2019). Most of this preparation is accomplished through the operation of hospitals and care facilities throughout the United States. Medical staff work and train in these hospitals, which allows them to maintain and improve medical skills during peacetime operations. For example, the DOD operates a certified Level 1 trauma center in San Antonio, Texas. This facility is authorized to provide trauma services for non-military patients in order to provide a platform for uniformed medical providers to maintain critical trauma skills (Sanchez, 2018).

The secondary purpose of the MHS is its *benefits mission*, which more closely resembles the missions of civilian healthcare systems. The peacetime mission of the MHS is to provide medical services to eligible beneficiaries, including retirees and family members of active and retired servicemembers. Services are provided to non-uniformed beneficiaries (e.g., family members, retirees) in military facilities on a space-available basis. This allows military providers to maintain their training and readiness for a diverse patient population in the military hospitals, medical centers, ambulatory care facilities, and dental clinics in the direct care system (Adirim, 2019).

Performance objectives in the MHS are guided by the four tenets of the MHS Quadruple Aim. The MHS Quadruple Aim is similar to the Institute for Healthcare Improvement's triple aim of improving the experience of care, improving the health of populations, and reducing per capita costs of healthcare (Berwick, Nolan, & Whittington, 2008). The key distinction for the MHS is the addition of a fourth aim—readiness. The MHS readiness aim encompasses accountability for "ensuring that the total military force is medically ready to deploy and that the medical force is ready to deliver healthcare anytime, anywhere in support of the full range of military operations, including humanitarian missions" (DHA, 2013).

Governance

The MHS executes its mission through five major DOD organizations: Office of the Assistant Secretary of Defense for Health Affairs, DHA, MEDCOM, Navy Bureau of Medicine and Surgery, and Air Force Medical Services (Mendez, 2018b). Leadership is comprised of a combination of uniformed personnel and civilian government employees (DHA, 2019d). Each organization executes distinct responsibilities, but they cooperate through a federated governance structure that enables collaboration, resource sharing, and streamlining of administrative functions (see Figure 1).

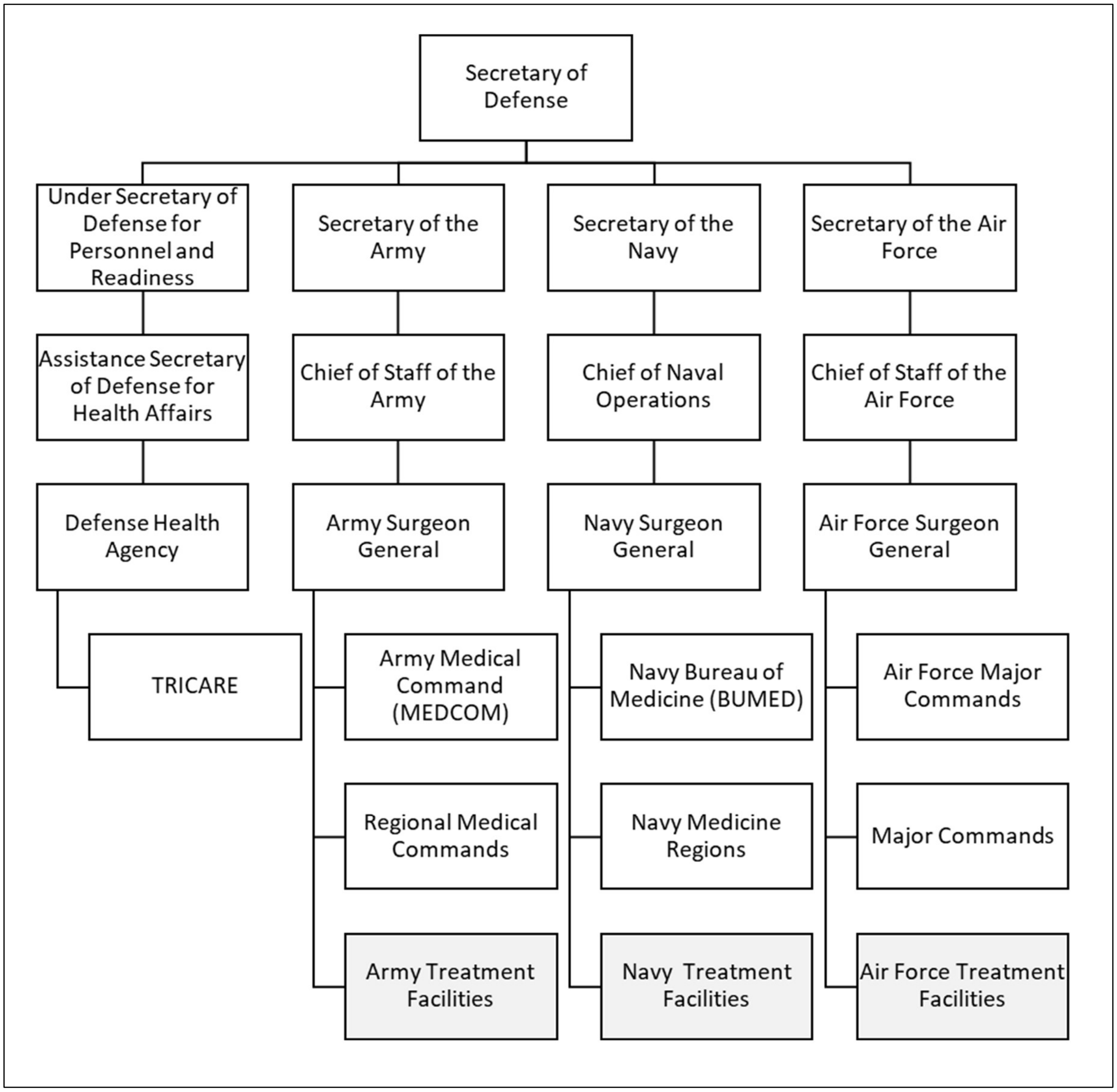

Figure 1. Governance structure of the MHS through October 1, 2021. This organization chart depicts the organization structure of the MHS prior to the transfer of management authority of all MTFs to the DHA. Adapted from "Military Medical Care: Frequently Asked Questions, Report No 45399, Version 4," by Bryce Mendez, 2018, Congressional Research Service, p. 3.

Prior to 2013, the operational aspects of the MHS were divided among the medical departments of the three uniformed services (Army; Navy, to include the Marine Corps; and Air Force) with each service independently operating its own clinics, hospitals, and medical centers (Bono, 2017; DOD, 2014). In September 2013, U.S. Department of Defense Directive 5136.13 established the DHA, which effectively unified these three subcomponents (DOD, 2013).

The role of the DHA is to manage the TRICARE Program; manage and execute the Defense Health Program (DHP) appropriation; support the coordinated management of multi-service markets (MSM);[1] and exercise management responsibility for shared services, functions, and activities of the MHS (Mendez, 2018a, p. 5). The DHA was created in order to reduce costs by consolidating services, providing better coordination of resources, promoting more rapid flow of information, and achieving more informed decision-making (Basu, 2012; Bono, 2017; Cain, 2013; Joint Chiefs of Staff [JCS] 2015). The DHA established common business practices and integrated functions among the three uniformed services. These integrated functions include purchasing medical supplies, management of the TRICARE health plan, pharmacy services, health information technology, education and training, contracting, facility planning, resource management, and research and development (Collins, 2015; Hutter et al., 2019; DOD, 2013). The DHA also increased interoperability among the services, especially with respect to technology and the management of access to care through the TRICARE health plan. As a result of cost-saving initiatives and efficiencies gained by creating the DHA, the rate of healthcare spending growth slowed throughout the MHS.

In 2017, Congress further expanded the role of the DHA with the passage of the FY 2017 NDAA. The NDAA is a series of federal laws that are passed each year to specify the budget and expenditures of the DOD. Aside from authorizing the annual budget, the NDAA also specifies how certain funds or activities within the DOD will be managed. The 2017 NDAA directs the MHS to consolidate the management of all DOD MTFs under the DHA. The purpose of this transformation is to eliminate duplicative activities, achieve efficiencies in management functions, and to reduce headquarters-level personnel requirements (GAO, 2018c).

[1] Multi-service markets are healthcare markets in which the clinics or hospitals from two or more uniformed services have overlapping service areas (DHA, 2019h; TRICARE, 2019).

As part of this centralization process, the DHA will assume responsibility for hospital budgets and common performance standards pertaining to readiness, quality, access, outcomes, and safety (Smith et al., 2017).

Transformation efforts commenced in 2018 and will continue through 2021. Congress requires the DOD to transfer 457 MTFs to the administration of the DHA by October 1, 2021 (NDAA, 2019). Figure 2 depicts the governance structure of the MHS once this change has been fully implemented.

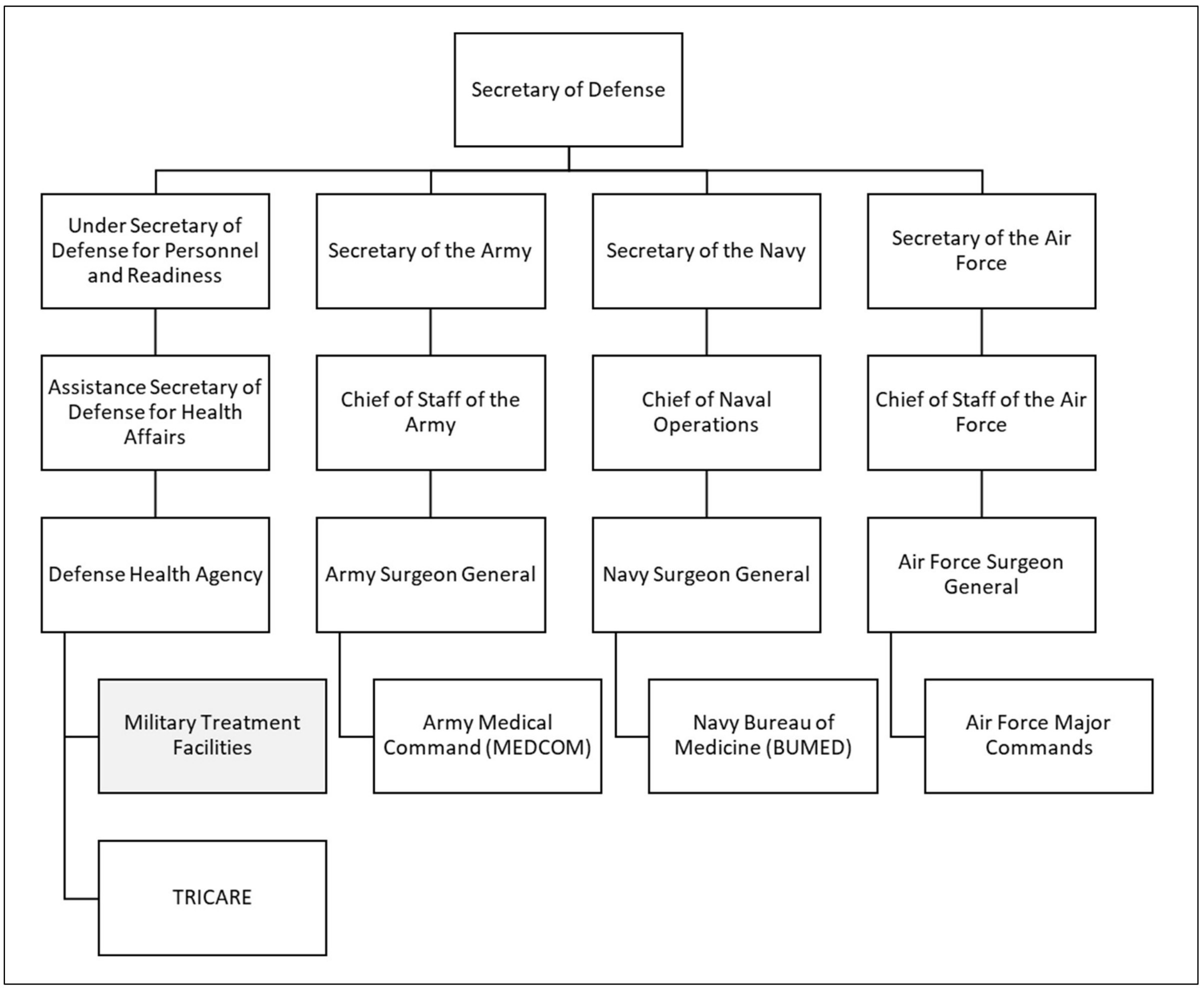

Figure 2. Governance structure of the MHS after October 1, 2021. This organization chart depicts the organization of the MHS after the DHA assumes management of MTFs after October 1, 2021. Adapted from "The U.S. Military Health System: Promoting Readiness and Providing Health Care, by Terri Tanielian and Carrie Farmer, 2019, *Health Affairs, 38*(8), p. 1260.

Funding

The MHS is funded through congressional appropriations in the *Unified Medical Budget*, which is managed by the Assistant Secretary of Defense for Health Affairs. This budget includes resourcing for medical personnel, military construction projects, operational costs of military medical facilities, and non-war-related health services (Mendez, 2018b). Funding for the delivery of health services is covered by the DHP, which typically constitutes approximately 67% of the total unified health budget (Mendez, 2018b). The unified budget request of $50.6 billion for FY 2019 includes $33.7 billion for the DHP, $8.9 billion for military personnel, $0.4 billion for military construction, and $7.5 billion for healthcare accrual (GAO, 2018a). This outlay encompasses approximately 9% of the total defense budget (DHA, 2018a, 2019e; Mendez, 2018b).

Once Congress appropriates funds to the Unified Medical Budget and DHP, they are disbursed to each of the three services (Army, Navy, and Air Force). The medical departments of each service in turn allocate funding to each MTF (see Figure 3). Almost all of each facility's funding comes from this budget process, though there are a few secondary funding sources. For example, 10 U.S.C. §1097(b) authorizes MTFs to undertake third-party collections to other insurance payers for beneficiary care that is delivered in treatment facilities if enrollees carry such extra insurance (Mendez, 2018b). Additionally, a standing authorization exists for transfers between the Medicare Eligible Retiree Health Care Fund and TRICARE that reimburses MTFs for the cost of healthcare delivered to Medicare-eligible military retirees in military facilities. Supplementary funding is also provided to MTFs through other reimbursable special programs and transfer authorities (Mendez, 2018b).

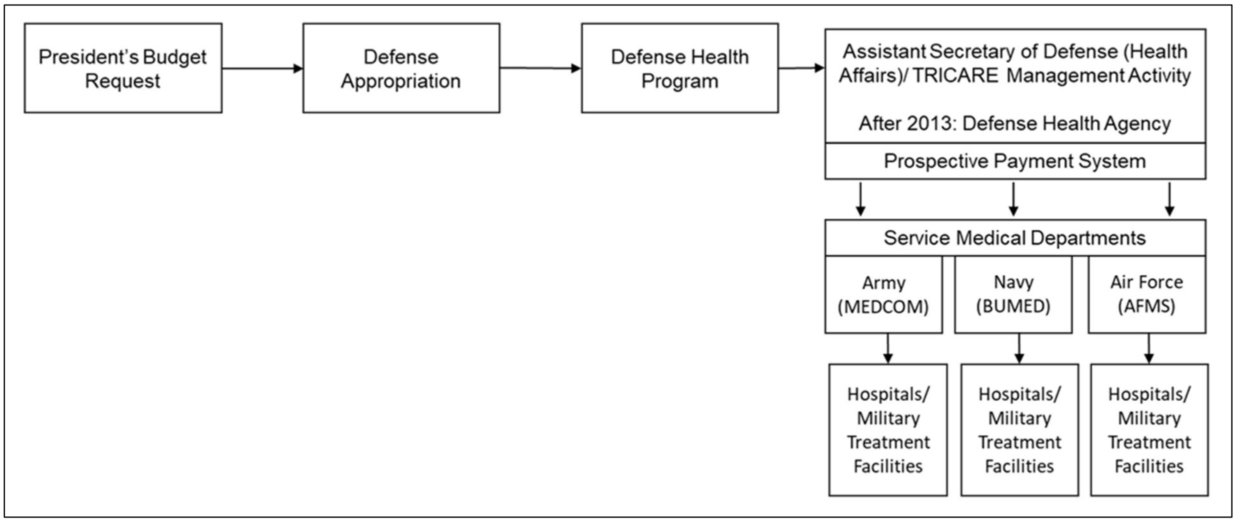

Figure 3. MHS funding process. Adapted from "Performance-Based Adjustment Model Overview" presentation by Robert Griffith on March 26, 2013 at the HFMA 2013 Texas State Conference.

Facilities

As of 2019, the direct care system is comprised of 51 inpatient hospitals and medical centers, and 672 ambulatory care, occupational health, and dental clinics (Adirim, 2019, p. 1269). These facilities, collectively referred to as MTFs, are typically located on military installations or in areas densely populated by military personnel and their families. They are operated in a federated manner by the medical departments of the Army, Navy, and Air Force, and overseen by the DHA (Adirim, 2019; Tanielian & Farmer, 2019).

MTFs are divided into three facility types. *Health clinics* are defined as ambulatory care medical facilities that do not provide inpatient services. Health clinics typically offer primary care services and a limited number of specialty care services (DHA, 2019g). *Hospitals* are defined as facilities that offer both inpatient and outpatient services and a variety of specialty care services. *Medical Centers* are defined as large hospitals that offer both inpatient and outpatient care, as well as a combination of specialty and subspecialty care services. Medical Centers typically operate Level 1 Trauma Centers that are authorized to provide emergency care for non-military patients in their respective communities. Due to the diversity of specialty care provided, Medical Centers also host Graduate Medical Education (GME) programs and medical research (DHA, 2019g).

Key Distinctions Between the MHS and the Private Sector

Several characteristics of the MHS distinguish it from other health systems in the United States. Terry Adirim (2019) noted, the MHS "is more akin to a single payer than to the fragmented civilian sector system [in the United States]" (p. 1270). Due to this distinction, the MHS is not bound by the same market forces that drive strategy in civilian healthcare organizations. Instead of attempting to gain market share and increase revenues, military organizations survive by seeking efficiencies that drive down costs to the federal government and increase value to taxpayers (Adirim, 2019). This is one of the drivers for consolidation efforts, including the creation of the DHA (GAO, 2018; Smith et al., 2017).

In addition to lowering costs, MHS organizations are also driven to avoid low-value healthcare services. Since MHS providers are not paid through fee-for-service arrangements, there is less incentive to order unnecessary diagnostic tests, conduct nonessential follow-up visits, or prescribe clinically inappropriate care such as antibiotics for viral infections. There is evidence that patterns of low-value care differ between the MHS and private sector counterparts. For example, a comparative study between MHS providers and purchased care (civilian) providers found that private sector providers prescribe expensive, low-value care more often than MHS providers (Koehlmoos et al., 2019). Another study on defensive medicine also found that patients in military healthcare facilities receive less intensive medical services with no measurable adverse impact on outcomes (Frakes & Gruber, 2019).

As discussed, the primary mission of the MHS also differs markedly from commercial counterparts. While most civilian healthcare organizations (e.g., for-profit and not-for profit community hospitals) provide healthcare services to general patient populations or to specific types of patients, such as children, veterans, women, or cancer patients (Bolon, 2005), military healthcare organizations have the added role of providing medical support for military operations. To support these unique requirements, MHS organizations often adopt strategies that would not be tenable in typical civilian healthcare facilities. For example, the DOD maintains hospitals in isolated areas with low populations. These are often located near some of the largest defense training areas such as the National Training Center in Fort Irwin, California; the Marine Corps Air Ground Combat Center in Twentynine Palms, California; and the Joint Readiness Training Center in Fort Polk, Louisiana. Maintaining hospitals in such deeply rural areas would not be financially viable for most civilian hospitals without a government or charity funding source, yet the DOD maintains these facilities because they are required to support high-risk training operations and to provide medical services to the remotely stationed servicemembers and families.

Another key distinction for military healthcare is the maintenance of critical services that are not available within the local healthcare markets. For example, many military hospitals maintain inpatient services despite a very low daily inpatient census. The average daily inpatient census across the MHS for FY 2016 was 37.4 with many facilities falling well below the minimal census required to keep the doors open in a civilian facility (Bono, 2017).

Some scholars have believed that these additional support requirements elevate the costs of military healthcare; however, these costs are difficult to quantify and have not historically been measured (Adirim, 2019, p. 1270). Though it may be difficult to conduct straightforward comparisons between military and civilian healthcare costs (GAO, 2018a), there are some requirements specific to military healthcare organizations that are easily measurable. For example, DOD (2016) regulations require every servicemember to undergo an annual physical assessment (DODI 6200.06), which drives up yearly healthcare utilization rates by minimum of one encounter per servicemember per year. Additional health assessments are also required before and after deployments, in addition to care provided for any deployment-related health conditions. These assessments come with measurable costs that are unique to military healthcare organizations.

Other financial impacts, such as the influence of military requirements on provider turnover and discontinuity of care, are less easily quantified and measured. The financial consequences of staff turnover in civilian healthcare organizations are well documented in the existing literature (Gray, Phillips, & Normand, 1996; C. B. Jones, 2005; Waldman, Kelly, Arora, & Smith, 2010). These effects may be exacerbated in MHS facilities because military-specific requirements for active duty medical staff (e.g., training, deployments and permanent reassignments) add to the organizational stressors associated with the typical drivers of turnover. Deployed staff are usually replaced by temporary backfills from the U.S. Reserves or contracted civilian providers, although it is common for facilities to experience periods of underlap during these transitions (GAO, 2018a).

In addition to deployments, military facilities are also impacted by the frequent reassignment of military servicemembers. On average, military servicemembers are reassigned to a different geographic location at least once every 36 months, although moves can be more frequent due to requirements for professional development, training, and deployments (Vergun, 2013). Frequent moves by active duty military members can also have a direct impact on non-military (civilian) employees in healthcare facilities. Many government service employees are family members of active duty personnel and move just as frequently as their military spouses.

Research has shown that these military-specific disruptions impact healthcare delivery in MHS facilities. For example, a study of military physicians preparing for deployment found that their patients experienced a 15-30% increase in specialist visits and a 15-18% increase in emergency room visits during the deployment and pre-deployment periods (Schwab, 2018). These changes were attributable to discontinuity of care as their physicians prepared for an extended departure. Another study of military hospitals from 2001 to 2006 found that hospital productivity was negatively impacted by support requirements for the wars in Iraq and Afghanistan (Childress, 2013). These studies highlight the fact that readiness mission of the MHS is linked to financial and performance consequences for hospitals that add to the complexity of comparisons to civilian healthcare organizations.

Cost, Quality, and Access in the MHS

Like many civilian healthcare organizations, cost, quality, and access are three of the biggest strategic concerns in the MHS. Cost has become increasingly important to Congressional, DOD, and MHS leaders over the last 20 years due to ballooning defense expenditures on healthcare. From 2001-2017, the MHS experienced a 217% increase in costs, prompting a call for intervention from the GAO (2017). From 2001 to 2010, spending on healthcare rose across the DOD by an average of 11.6% annually, which was significantly higher than the average 6.4% increase in healthcare spending across the United States for the same time period (Adirim, 2019; CMS, 2019c). Throughout the last 20 years, healthcare costs have consumed an average of 8-10% of the total annual DOD budget (Mendez, 2018a).

These increases are attributable to several factors. The wars in Iraq and Afghanistan created an increased demand for healthcare services among servicemembers. Additionally, the costs associated with resourcing an agile medical force to support combat operations also places a strain on healthcare funding. Another factor is the expansion of healthcare benefits in the TRICARE program. Starting in 2001, Congress initiated the TRICARE for Life program that provided secondary insurance coverage for over 2 million Medicare-eligible military retirees and their dependents (DHA, 2018b). A generous new pharmacy benefit also expanded prescription drug coverage for this same population of Medicare-eligible military retirees (Adirim, 2019; DHA, 2018b). Though these new benefits are resourced through allocations of money set aside for retiree healthcare, they are often included in estimates of military healthcare spending.

Healthcare quality in the MHS is another critical concern for Congress, DOD, and MHS leaders. As noted, active duty servicemembers are required to maintain a high level of physical health (i.e., medical readiness) in order to perform military duties in demanding environments. The healthcare services that ill or injured soldiers receive have the potential to impact the amount of time they are available to perform their military duties. Since poor quality healthcare can impact the overall readiness of the U.S. military force, the quality of military healthcare has been identified as a national defense priority (Pellerin, 2017).

Congress has closely monitored healthcare quality in the MHS over the last 20 years. Media attention regarding conditions at the Walter Reed Hospital (NPR.org, 2011; Priest & Hull, 2007), allegations of medical errors (LaFraniere & Lehran, 2014a, 2014b), complaints about access to care (Kime, 2014), and organizational changes across the MHS prompted the Secretary of Defense to order a 90-day review of quality, access, and safety across the MHS in

2014. This comprehensive study, known as the MHS Review, provides the basis for a considerable amount of literature that has been published on military healthcare in the last 5 years. The MHS Review commissioned a panel of military and non-DOD healthcare experts who compared the performance of the MHS to civilian counterparts using nationally recognized performance benchmarks. Overall, the MHS Review found that the MHS provides high quality, safe, and timely care that is comparable to the civilian sector. However, the review also noted wide performance variability, with the MHS performing better than national benchmarks in some areas and below national benchmarks in others (DOD, 2014). This was one of many factors that laid the foundation for ongoing quality improvement efforts across the MHS. These efforts emphasize reducing variation and transitioning into high reliability organizations with respect to safety and quality.

Access to care is another major concern for MHS leaders. The IOM (1993) defined *access to care* as “the timely use of personal healthcare services to achieve the best health outcomes” (p. 4). Although no universal definition of timely access to care has been established nationally, the access standards for the MHS are codified by law (32 C.F.R. §199.17(p)(5) - TRICARE program, n.d.). Access standards for military beneficiaries include (32 C.F.R. §199.17(p)(5) - TRICARE program, n.d.; DOD, 2014):

1. No more than 30-minute drive time for primary care
2. Specialty care appointments available within 4 weeks
3. Routine medical appointments available within 1 week
4. Wait time for urgent care appointments will not exceed 24 hours
5. Emergency room access available 24 hours per day/7 days per week
6. No more than 60-minute drive time for specialty care appointments

7. Office wait times not to exceed 30 minutes unless emergency care is being rendered to another patient

These access standards are met through a combination of healthcare services provided in the direct and purchased care systems. The purchased care component serves as a safety net when access to care standards cannot be met through the direct care system or when military facilities are over capacity (32 C.F.R. §199.17(p)(5) - TRICARE program, n.d.; DOD, 2014). Access to care is monitored and managed at each MTF through designated Group Practice Managers (Air Force) or Access to Care managers (DHA, Army, and Navy; DOD, 2014). The role of these access managers is to monitor access, communicate with clinic leaders, and coordinate care through the purchased care system when access standards cannot be met in military facilities.

The 2014 MHS Review found that, on average, the MHS meets or exceeds its own access to care benchmarks (DOD, 2014). According to this review, the average wait time for specialty care appointments in the direct care system is 12.4 days, while the average wait for urgent (non-emergency) appointments is less than 24 hours in most facilities. However, the MHS Review panel also noted that there is high variability among MTFs regarding compliance with access-to-care mandates (DOD, 2014). The cause of this variability has not been studied empirically, but operational requirements for military medical staff may be one factor. Short-term deployments of military providers often disproportionately affect hospitals located in close proximity to deploying combat units and are often difficult for MTF leaders to program into long-term manpower planning. According to the GAO (2018a), MTFs often struggle to fill positions vacated by deployed medical staff due to the length of the federal and civilian hiring process, uncompetitive wages for federal civilians and contractors, and federal hiring freezes.

When MTFs are unable to meet access to care requirements in the direct care system, they rely on the purchased care system to supplement access. Since participation in the TRICARE program is voluntary for contracted civilian providers, this safety net does not always guarantee that there will be enough participation among local providers to meet access to care standards with purchased care. Studies have shown that the acceptance rate for TRICARE insurance among primary care providers (67%) is less than acceptance rate for private insurance (95%) and Medicare (86%) and is comparable to Medicaid (65%; Ben-Shalom, Schone, & Bannick, 2019; Boukus, Cassil, & O'Malley, 2009). These low acceptance rates have caused TRICARE beneficiaries to report problems with access to care in select markets (Ben-Shalom et al., 2019; GAO, 2006, 2011, 2013).

In addition to monitoring access to care using objective metrics such as the average wait time for appointments, the MHS also monitors access to care using subjective measures such as satisfaction surveys. Assessment of these two types of measures often produce conflicting findings in relation to access. For example, the MHS Review noted a discrepancy between documented MTF compliance with access-to-care standards and patients' reported satisfaction with timely access to care (DOD, 2014). Another study found that TRICARE-insured family members report access to care that is similar to uninsured or publicly insured populations (Seshadri, Strane, Matone, Ruedisueli, & Rubin, 2019). The cause of these discrepancies between self-reported access and objective measures of access has not yet been empirically determined, although low reimbursement rates and low acceptance rates for TRICARE insurance in the purchased care component may be one factor (Ben-Shalom et al., 2019; GAO, 2006, 2011, 2013).

Since Congress and MHS leaders are accountable to the American public for the value of military healthcare services delivered through taxpayer funding, matters related to cost and quality have generated congressional attention over the last 20 years. Many of these concerns prompted lawmakers and military leaders to seek policies aimed at reducing costs and increasing healthcare quality. This provided the impetus for the PBB programs that the MEDCOM has experimented with over the past 15 years. Though the Army's PBB programs included performance measures for cost, quality and access, this study focuses specifically on the quality measures.

Summary

The unique demands of the MHS required a tailored approach for addressing performance that differed from the traditional pay for performance approaches that were popular among civilian hospitals in the 2000s and beyond. Because the MHS operates with a federated governance structure and funding resources are shared between facilities, traditional pay for performance approaches are less appropriate in this context. MHS healthcare facilities do not have financial relationships with multiple insurers, nor do they operate with a traditional business model in which more money can be earned through greater efficiency, productivity, marketing, or patient volume. Instead, PBB is a better approach because healthcare facilities are resourced through a finite federal budget that is allocated through a top-down approach. Performance improvement can be encouraged in this context by including performance thresholds in the budgeting process for each facility, such that higher performing facilities are rewarded with greater funding. Though this study does not provide a direct comparison to pay for performance, it examines PBB to determine if it has similar impacts on quality performance that are observed with traditional pay for performance programs.

Evolution of PBB in U.S. Army Healthcare Facilities

In 2004, a select group of Army hospitals in the southeast region of the United States were chosen to participate in pilot program testing PBB. There were several drivers for this transition, but the main catalyst for the early program was cost control. As described above, mounting fiscal pressures from increased healthcare costs and the impacts of the dual wars in Iraq and Afghanistan prompted Army leaders to seek alternatives to the traditional method of building treatment facility budgets. In light of these fiscal pressures, MEDCOM initiated sweeping changes to its funding model for all of its facilities starting in 2006. This program, known as the Performance-Based Adjustment Model (PBAM), was a mandatory Army-wide program intended to address concerns over unsustainable healthcare costs, improve accountability, and improve the quality of healthcare services (Griffith, 2013; West et al., 2010) by making a portion of MTF funding contingent upon performance (i.e., PBB).

Prior to 2006, Army hospitals were funded in the traditional manner: Congress appropriated funds to the DHP, which were then disbursed to each of the three services and allocated to the MTFs through the services' medical departments. Under this arrangement, Army MTFs typically received "last year's budget plus inflation" (West & Cronk, 2011, p. 32) at the beginning of each FY. This allocation method is associated with some undesirable effects (Balakrishnan, Soderstrom, & West, 2007; West et al., 2010). Many scholars have argued that the annual U.S. fiscal cycle does little to support performance and long-term investment. *Budget lapsing* refers to a typical feature of government programs in which unspent funds do not carry over from one budget cycle to the next (Balakrishnan et al., 2007; Zimmerman, 2003). The threat of having unspent funds removed from future budgets or transferred to other organizations creates a disincentive for public managers to improve efficiency or implement cost-saving techniques (McNab & Melese, 2003). As a result, managers often engage in a counterproductive method of saving and dissaving throughout the FY.

Balakrishnan et al. (2007) documented this spending pattern in a 5-year study of 31 Army hospitals. They found evidence that administrators were managing uncertainty by maintaining a reserve of funds throughout the year, only to be expended in a large outflow at the end of the FY. Not only was this funding method and its resultant spending patterns inefficient, it did little to incentivize more constructive pursuits, such as investment in quality improvement or enhanced productivity. West et al. (2010) argued that the traditional funding model did little to hold administrators accountable for the output of their organizations. When administrators and managers expect their funding levels to remain relatively stable regardless of their performance or productivity, there is little incentive to increase output or improve quality. Because the primary performance indicator for administrators and managers is staying within budget, there is little fiscal incentive to increase output or improve quality under the traditional funding model.

The PBAM was designed to address these concerns by "aligning hospital- and department-level managers' incentives with both funding and MEDCOM's strategic goals" (West et al., 2010, p. 53). PBAM reversed the existing incentive structure by rewarding managers for efficiency, productivity, and quality improvement rather than penalizing departments for returning unused funds. Even though MEDCOM could not change the way that funds were disbursed from Congress and the DOD, PBAM allowed high-level leaders to adjust how the funds were disbursed across the system (West et al., 2010, p. 53). This enabled MEDCOM to implement a new incentive structure that incorporated performance information in budget decisions across Army hospitals.

Based on the outputs versus outcomes paradigm of PBB that was previously described (M. S. DeVries & Nemec, 2019), the PBAM program can be categorized as a hybrid of both of these goals. It was designed to resource hospitals based on a combination of productivity measures (outputs) and quality measures (outcomes; see Figure 4).

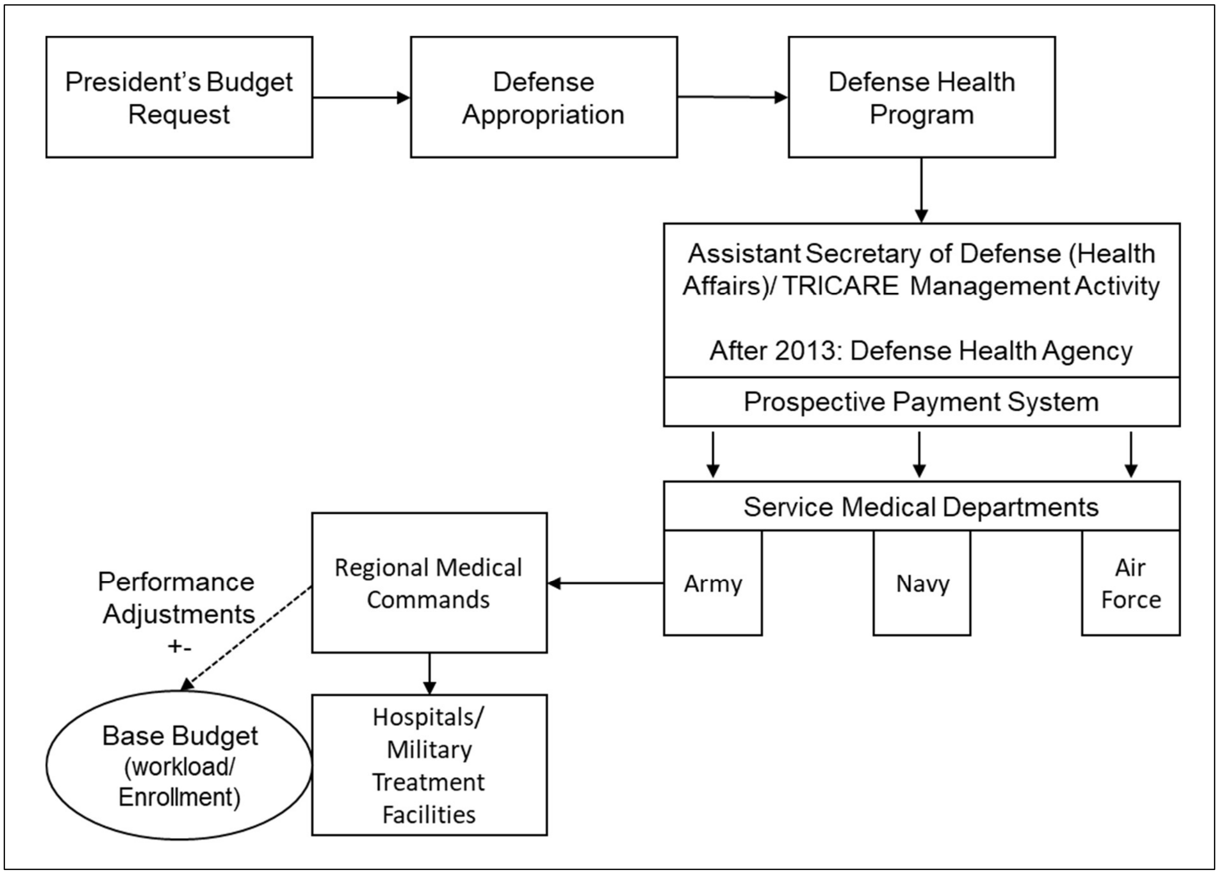

Figure 4. Performance-based resourcing of U.S. Army hospitals under PBAM and IRIS programs. Adapted from "Performance-Based Adjustment Model Overview" presentation by Robert Griffith on March 26, 2013 at the HFMA 2013 Texas State Conference.

Under the PBAM program, MEDCOM resourced MTFs at the beginning of the FY based on productivity data that was determined by workload and efficiency measures (Rheney, 2007). Throughout the year, adjustments (decrements and bonuses) were applied to facility-level funding based on performance in three areas: capacity (workload and efficiency), administrative performance (i.e., coding accuracy, cost accounting, staffing management) and quality (i.e., evidence-based practice, patient satisfaction). Though performance adjustments were calculated and reported monthly, actual budget adjustments typically occurred three to four times per year (Griffith, 2013, 2019). This encouraged administrators to closely monitor performance on a monthly basis in order to forecast funding levels throughout the year.

PBAM performance adjustments represented a small portion of the overall MTF budget; no more than 5% of the total budget was *at risk* for adjustment. On average, actual bonuses (and penalties) typically added up to approximately 3% of the budget (Griffith, 2019). Despite the small size of the incentive, there was initial evidence suggesting a link to increased performance. A pilot study conducted with six Army healthcare facilities in the southeast region of the United States provided an initial proof of concept. This model was then endorsed by the Army Surgeon General and rolled out to the rest of the Army in 2006. After implementing PBAM, Army facilities across MEDCOM realized a 13.9% increase in total output, a 3.3% gain in provider efficiency, and a 15% gain in coding accuracy (West et al., 2010). Initial gains were also documented for quality measures such as patient satisfaction and population health screening (Griffith, 2013; Waymire & West, 2010). For example, prior to 2006, Army facilities significantly underperformed on measures in the HEDIS relative to Air Force and Navy peers. By 2009, this trend reversed and Army outperformed Navy and Air Force facilities on the same measures (Griffith, 2013).

In 2014, the U.S. Army transitioned from PBAM to another PBB program known as the Integrated Resource and Incentive System (IRIS). The purpose of IRIS was to “align funding & incentives in order to enhance MTF value production” (Dunning, 2014, Slide 6). This program was very similar to PBAM, except that it integrated the workload-based resourcing system with the performance-based incentive system (Griffith, 2019). The key distinction between the IRIS and PBAM programs was the introduction of a subcapitation model for primary care at each facility.

Under the new IRIS model, Army hospitals were funded based a *per member per month* capitated rate that was directly linked to the number of patients enrolled to their facilities. Funding levels at each facility were projected at the beginning of each FY based on the anticipated enrollment of patients at that facility. Each facility had the opportunity for base funding levels to increase or decrease throughout the year based on enrollment changes and performance on pre-established metrics. The IRIS program promoted even greater involvement of administrators and managers at each facility because leaders were required to submit a performance plan at the beginning of each FY. These changes also required leaders to focus on enrollment rather than RVU production alone in order to increase base funding levels (Dunning, 2014). This new structure emphasized high-value healthcare services that were intended to keep patients healthier and reduce the need for future preventable healthcare services. This strategy also reduced the incentive to deliver low-value healthcare services that were not evidence-based.

The incentive structure for quality was essentially the same for both the PBAM and IRIS programs. Although the incentive size and the number of performance measures shifted from year to year, both programs rewarded quality by incorporating small, organization-level budget adjustments (bonuses or penalties) based on performance over a rolling 12-month period. Under both programs, a small portion of each facility's overall budget was placed "at risk" contingent upon quality performance. From 2006-2018, quality metrics were added, dropped and revised annually for both the IRIS and PBAM programs (see Table 1 for the yearly evolution of metrics from 2006-2018), giving Army leaders the ability to align incentives with changing strategic priorities (Griffith, 2013, 2019).

Table 1

Incentivized Quality Measures Under PBAM and IRIS, 2004-2019

Measure	Year	Measure source	Definition
Breast cancer screening	2004/2006*	NCQA/HEDIS	Percent of eligible enrollees (age 42-69) with screening mammography within the last 24 months.
Cervical cancer screening	2004/2006*	NCQA/HEDIS	Percent of eligible enrollees (age 24-64) with an appropriate cervical cancer screening in the past 36 months.
Diabetes-A1C Screening	2004/2006*	NCQA/HEDIS	Percentage of eligible patients with diabetes who have had A1C testing within the past 12 months.
Diabetes-A1C Control	2004/2006*	NCQA/HEDIS	Percent of eligible patients with diabetes with HbA1C levels at nine or below within the last year.
Diabetes-LDL Control	2004/2006*	NCQA/HEDIS	Percent of eligible patients diagnosed with diabetes who have an LDL level below 100mg/dl within the last year.
Asthma Care	2004/2006*	NCQA/HEDIS	Percent of eligible enrollees (age 5-56) with persistent asthma who are prescribed medications considered acceptable as a primary therapy for the long-term control of asthma.
Colorectal cancer screening	2006	NCQA/HEDIS	Percent of eligible enrollees, age 50-80, who have had appropriate colorectal cancer screening. Screening intervals vary according to the method of screening.
Chlamydia screening	2009	NCQA/HEDIS	Percent of enrolled active duty women, age 16-25, who have received a chlamydia screening within past 12 months.
Pneumonia vaccination status	2009	Managed Care Organization Accreditation Benchmarks and Thresholds	Percent of eligible enrollees 65 years older who have received one or more pneumococcal vaccinations.
Patient satisfaction (APLSS)	2009	Army MEDCOM	Percent of patient satisfaction surveys returned with a rating of 4 or 5 (out of 5) on Question 21, the “overall satisfaction” item.
Inpatient professional service rounds	2009	Army MEDCOM	Number of physician-associated inpatient episodes of care meeting documentation standards (based on E & M codes).

Measure	Year	Measure source	Definition
Antibiotics w/in 1 hour of surgery	2010	Joint Commission/ ORYX	Number of surgical patients with prophylactic antibiotics initiated within 1 hour prior to surgical incision (2 hours if receiving vancomycin).
Children's asthma care: Home management plan of care	2010	Joint Commission/ ORYX	Percentage of cases that have received the appropriate treatment identified to provide the best outcome for the patient.
Patient satisfaction w/access	2010	DHA	Number of surveys returned with a 4 or 5 rating (out of 5) on Questions 9, 11, and 13 of the APLSS Survey (items pertaining to satisfaction with access).
PCM continuity	2011	DHA	Number of primary care encounters in which the patient saw his or her assigned Primary Care Manager.
Documented Body Mass Index in health record	2011	DHA	The percentage of enrollees with a calculable BMI in medical record.
Well child visits (6 in 15 mos.)	2012	NCQA/HEDIS	HEDIS criteria for percent of eligible population with six or more well child visits in the first 15 months of life.
Inpatient satisfaction	2012	DHA	Percentage of patients satisfied with MTF provided care for TRISS Question 20.
Preventable admissions	2014	AHRQ	Number of admissions meeting AHRQ definitions for "potentially preventable."
Readmissions	2014	CMS/ Partnership for Patients	Penalty per Re-Admission for Three Focus Areas: Acute myocardial infarction, heart failure, and pneumonia. Date of readmission, for the same diagnosis is within 30 days of discharge date.
Tobacco use	2014	DHA	The number of enrollees without tobacco use in the previous 12 months.
Healthy weight	2014	DHA	Number of enrollees 24 months and older that had a Healthy Weight measured in a Primary Care Clinic within the prior 12 months using CDC standards.
Tobacco use	2015	DHA	Number of enrollees 18 years or older who report that they do not use tobacco.
Low back pain	2016	NCQA/HEDIS	Percent of enrollees aged 18-50 with a primary diagnosis of low back pain who did not have an imaging study (plan x-ray, MRI, CT scan) within 28 days of diagnosis.

Measure	Year	Measure source	Definition
PTSD depression treatment dosage	2016	DHA	Number of patients initially diagnosed with PTSD or Major Depressive Disorder who met the requirement of initial visit plus three or more follow-up visits within the first 90 days after diagnosis.
PTSD depression treatment outcome	2016	DHA	Number of PTSD and depression episodes of care with either a clinical response or full remission.
Exclusive breastfeeding	2016	Joint Commission	Overall rate of newborns that were exclusively fed breast milk during the entire hospitalization.
Body Mass Index outcome	2016	DHA	Number over enrollees with a BMI > 25 who have at least three primary care visits and achieve 1.0% BMI reduction per person.
All cases morbidity index	2017	(NSQIP) / American College of Surgeons	Measures the quality of surgical care using a case-mix adjusted, risk-adjusted, and outcomes-based odds ratio to compare actual performance to expected performance.
CLABSI	2017	CDC NHSN	Risk-adjusted rate ratio of CLABSI in MTF ICUs compared to other participating ICUs in the CDC NHSN Program.
Behavioral health outcomes	2017	DHA	Percent of patients with clinically significant improvement in three areas: depression, PTSD, anxiety disorder. Note: Extension of 2016 outcome metric.
Behavioral health provider data input	2017	DHA	Rate of provider input into the Behavioral Health Data Portal.
Medical management for high utilizers	2017	DHA	Number of high utilizing patients (10 or more primary care encounters in 12 months) receiving appropriate case management.
Foreign body retention	2017	DHA	Number of retained object events. A retained object is defined as a surgical object that is accidently left in the patient during a procedure.
Vaginal delivery with shoulder dystocia rate	2017	National Perinatal Information Center	Vaginal delivery with shoulder dystocia rate - linked to inborn ≥ 2500g with birth trauma.
Postpartum hemorrhage rate	2017	National Perinatal Information Center	Rate of women who experienced postpartum hemorrhage ≥ 1,000 ml.

Measure	Year	Measure source	Definition
Appropriate testing for children with pharyngitis	2017	NCQA/HEDIS	Assesses children 2-18 years of age who were diagnosed with pharyngitis, dispensed an antibiotic, and received a group A streptococcus test for the episode. A higher rate represents better performance (i.e., appropriate testing).
Appropriate treatment for children with upper-respiratory infection (URI)	2017	NCQA/HEDIS	Children 3 months to 18 years of age who were given a diagnosis of URI and were not dispensed an antibiotic prescription. A higher rate indicates appropriate treatment of children with URI (i.e., the proportion for whom antibiotics were not prescribed).
Follow-up after hospitalization for mental illness	2017	NCQA/HEDIS	Adult and children 6 years of age and older who were hospitalized for treatment of selected mental health disorders and received follow-up within 7 days of discharge and within 30 days of discharge.
Patient Satisfaction-Joint Outpatient Experience Survey	2017	DHA	Patient satisfaction incentive transitioned from APLSS to JOES survey. Question 22, overall satisfaction, was incentivized.

Note. LDL = low density lipoproteins. Information compiled from PBAM and IRIS Handbooks, years 2006-2017, with supplements from briefing slides 2004-2017.
[1] Some Army facilities participated in a pilot study of PBAM and started to be incentivized in 2004. Non-pilot sites started these incentives when PBAM was rolled out to all Army facilities in 2006.

The fee-for-service payments were intended to cover medical costs for patients not enrolled to the facility (such as servicemembers on short-term training assignments or in transition between permanent assignments). The supplemental funds covered the costs of designated special programs. The proportion of each facility's budget affected by supplemental funds varied from year to year based on several factors, such as the strategic priorities of the Army, and initiation of congressionally funded special projects. Due to the annual variation in special programs, enrollment, and funding at each facility, a full economic analysis of PBB in Army healthcare facilities is beyond the scope of this analysis.

To date, researchers have not conducted any rigorous studies to evaluate the effects of the Army's PBB programs on quality improvement. Analysts have conducted a small number of unpublished internal studies on performance improvement, but these were small in scope and did not include a comparison group. One unpublished study on the PBAM program did include a comparison to Navy and Air Force, but this study compared service-level performance and did not provide analysis to determine if service-level pre/post differences in performance were statistically significant. It remains unclear what effects the incentives in PBAM and IRIS had on the quality improvement in Army hospitals, if any.

Summary

This chapter began with an overview of the history, definitions, and research on pay for performance. This literature review demonstrates that the drive to improve the safety and quality of healthcare services in the United States has increased over the past 20 years. Pay for performance emerged over that period, but the research evidence supporting its use has yielded mixed findings. Some researchers have hypothesized that this heterogeneity in results may be due to patient factors, differences in care settings, and heterogeneity in program design features. Research findings for each of these factors was discussed. Overall, the literature on pay for performance suggests that it usually does not result in significant performance improvement. When performance does improve in conjunction with pay for performance programs, it is usually for process measures rather than patient outcomes, and those improvements tend to diminish or disappear over time.

This chapter also provided an overview of PBB and presented arguments for why it should be considered a *special case* of pay for performance that pertains to government programs. Rigorous research on PBB is limited, but the existing literature suggests that programs often fail to achieve their intended results. The most commonly cited reason for the failure of PBB programs is an inability to definitively link program performance to budget decisions. Unlike private sector organizations whose performance is primarily driven by market forces, public program resourcing is determined through political processes that may attenuate the link between program performance and program funding.

After PBB and its links to pay for performance, this chapter transitioned into an overview of the U.S. Army's PBB program in its healthcare facilities. Background information about the MHS was provided to establish the context for the emergence of the Army's PBB programs. Though PBB has been widely implemented in U.S. Army healthcare facilities, there is no experimental or comparative research to determine if it works in the Army or in other contexts.

Chapter 3: Theoretical Framework

Introduction

In this chapter, the basic tenets of RDT are reviewed. RDT concepts are then discussed in terms of their applicability to the MHS and the Army's adoption of PBB programs. Next, a conceptual model is presented that demonstrates how RDT can be used to predict the impact of PBB on quality performance in Army hospitals. The final section of this chapter presents hypotheses using the conceptual model and RDT as its basis.

Conceptual Framework

This study employs the conceptual framework of RDT (Pfeffer & Salancik, 1978) to examine the impact of PBB on quality metrics in U.S. Army medical facilities. The RDT framework has been used extensively in healthcare research related to financial incentives (Yeager, Menachemi, et al., 2014) and can therefore be useful in guiding this study on the effects of PBB incentives in military healthcare facilities.

According to RDT, organizations depend on critical resources from the external environment to function and survive. Organizations must develop strategies to manage or reduce dependencies on those external resources critical to survival (Pfeffer & Salancik, 1978). According to RDT, external forces in the environment can influence the strategies that organizations adopt in order to manage access to critical resources (Pfeffer & Salancik, 1978). Three environmental factors affect organizations' access to critical resources: munificence, complexity, and dynamism. *Munificence* refers to the overall supply and accessibility of resources in the environment (Pfeffer & Salancik, 1978). Environmental *complexity* and *dynamism* are both related to the overall degree of uncertainty in the market (Pfeffer & Salancik, 1978; Yeager, Savage, Ginter, & Beitsch, 2014). *Complexity* refers to factors that make strategic decision-making and actions more difficult (Pfeffer & Salancik, 1978). *Dynamism* refers to the level of change internal and external to the organization (Pfeffer & Salancik, 1978).

Most health services research studies that used RDT have examined private sector healthcare organizations such as hospitals, long-term care facilities (nursing homes and nursing subacute facilities), and medical practices (Yeager, Savage, et al., 2014). Though DOD healthcare organizations do not face the same environmental market pressures as private sector healthcare organizations, they still rely on resources in their environment to survive. Due to this dependency, the predicted relationships between organizations and their resource environment are also applicable to DOD healthcare organizations. The following section discusses key concepts in RDT and outlines how those concepts can be applied to the MHS resource environment to predict changes in hospital performance.

Application of RDT Concepts to the MHS

Healthcare studies applying the RDT framework define critical resources in many ways, although financial resources are the most common (Fareed & Mick, 2011). Other types of resources include reputation, prestige, status, and knowledge (Fareed & Mick, 2011). For the present study, the *critical resource* is financial. It is operationalized as the overall DHP budget allocation for military healthcare facilities, which is comprised of the portion of the U.S. National Defense budget that is set aside to cover the costs associated with delivering *in-house* healthcare within military health facilities (Mendez, 2018a).

Most healthcare studies investigate the impact of elements in the external resource environment on a dependent variable related to performance or organizational strategy (Yeager, Savage, et al., 2014). Examples of organizational strategies in the healthcare literature include participation in managed care (Yeager, Savage, et al., 2014; Zinn, Proenca, & Rosko, 1997), adoption of electronic medical records (Kazley & Ozcan, 2007; Yeager, Savage, et al., 2014), admission of obese nursing home residents (Zhang, Li, & Temkin-Greener, 2013), participation in the subacute care market (Weech-Maldonado, Qaseem, & Mkanta, 2009), adoption service or quality innovations (Banaszak-Holl, Zinn, & Mor, 1996; Fareed & Mick, 2011; Zinn, Weech, & Brannon, 1998), or some dimension of organizational performance (Hsieh, Clement, & Bazzoli, 2010; Yeager, Savage, et al., 2014; Zinn, 1994). For U.S. Army healthcare facilities that participated in the PBAM (2006-2013) and IRIS (2014-present) PBB programs, the *organizational strategies* refer to the performance improvement actions that Army healthcare facilities undertake to obtain bonuses or avoid penalties.

Pfeffer and Salancik (1978) described munificence in terms of the availability of critical resources in the environment. Limitations on critical resources in the environment can create uncertainty, which shapes the behavior of organizations in response to resource austerity (Fareed & Mick, 2011; Pfeffer & Salancik, 1978). In the context of PBB, fiscal climate may impact the degree to which performance measurement is used for resource decisions (Jordan & Hackbart, 1999; Lauth, 1985; U.S. General Accounting Office, 1993). In periods of fiscal stress, across-the-board cuts are more likely to be used regardless of program performance (Jordan & Hackbart, 1999; U.S. General Accounting Office, 1993). This reduces the incentive for organizations to improve performance and also reduces the availability of resources for organizations to devote to performance improvement efforts. In the context of military healthcare, a significant determinant of environmental munificence is the size of the DHP budget. The DHP budget impacts all military healthcare facilities and can fluctuate annually based on Congressional allocation through the NDAA (Mendez, 2018a; U.S. Congress, House Committee on Armed Services, 2017; DOD, 2018a). The DHP budget allocation shapes each medical facility's fiscal environment and affects the financial resources that are available for all operations, including quality improvement activities. During periods of fiscal austerity, facilities may have fewer resources available for performance improvement.

Any factor that makes a situation more intricate or complicated increases its level of complexity. In health services research, complexity has been operationalized as market competition among healthcare organizations (Banaszak-Holl et al., 1996; Hsieh et al., 2010; Kazley & Ozcan, 2007; Weech-Maldonado et al., 2009; Zinn et al., 1997; Zinn et al., 1998) because market competition increases the information uncertainty that decision-makers use to select organizational strategy in response to environmental variables (Yeager, Menachemi, et al., 2014, p. 52). The present study defines complexity in terms of patient demographic mix and acuity because these variables have the potential to complicate quality performance efforts.

For example, a facility with a higher population of elderly patients may be at greater risk for adverse health outcomes that can affect a facility's quality performance (Green, Passman, & Wintfield, 1991; Iezzoni, 2003). Patient age has also been found to be a predictor of satisfaction on military-sponsored patient satisfaction surveys (Mangelsdorff & Finstuen, 2003). Thus, facilities with a younger patient population may face more difficulty in achieving higher patient satisfaction score relative to peers. Complexity of healthcare needs in the patient population is another factor that may vary from facility to facility and impact organization-level quality performance. For example, during the peak periods of combat operations in Iraq and

Afghanistan, sick and wounded servicemembers were routed through four major facilities: Walter Reed Army Medical Center, Bethesda Naval Medical Center, Brook Army Medical Center, and Landstuhl Regional Medical Center (DHA, Wounded Warrior Care Center, n.d.; U.S. Army, 2012; U.S. Army Medical Department, 2016). This concentration of complex patients may make it more difficult for certain facilities to achieve higher quality performance relative to peer facilities with less vulnerable patients. Since patient safety is at increased risk during care transitions (Coleman, Parry, Chalmers, & Min, 2006; Forster, Murff, Peterson, Gandhi, & Bates, 2003), facilities serving higher acuity patients may be more vulnerable to adverse outcomes than facilities serving less complex patients requiring fewer care transitions.

Dynamism relates to the level of change in the resource environment. Factors that create fluctuations in the resource environment increase uncertainty and make it more difficult to predict the future and execute an organizational strategy (Yeager, Menachemi, et al., 2014, p. 52). The most common operationalization for dynamism in health services research is the unemployment rate (Yeager, Menachemi, et al., 2014), but it has also been operationalized in terms of number of managed care contracts (Menachemi, Mazurenko, Kazley, Diana, & Ford, 2012; Menachemi, Shin, Ford, & Yu, 2011) and proportion of Medicare and Medicaid inpatient days (Hsieh et al., 2010).

In the context of the current study, there are two major sources of dynamism: staff discontinuities and government shutdowns. Staff discontinuities in military facilities are common due to the dynamics of military deployments, military training exercises, and mandated movement of military servicemembers. When military staff members within hospitals deploy, it can impact staffing levels and create a disruption in continuity of medical services for patients (Schwab, 2018). Deployed staff are typically replaced by temporary backfills from the U.S. Reserves, but it is common for facilities to experience periods of underlap during these transitions. In addition to deployments, military facilities are also impacted by the frequent reassignment of military servicemembers. On average, military servicemembers are reassigned to a different geographic location at least once every 36 months, although moves can be more frequent due to requirements for professional development, training, and deployments (Vergun, 2013). Frequent moves by active duty military members can also have a direct impact on non-military (civilian) employees in healthcare facilities. Many government service employees are family members of active duty personnel and move just as frequently as their military spouses. High staff turnover and discontinuities can make it more difficult for organizations to execute a consistent performance improvement initiative. It can also cause disruptions in the patient-provider relationship, which can impact patient care (Schwab, 2018) and potentially impact quality performance.

Government shutdowns are another source of dynamism in the military healthcare environment. Government shutdowns contribute to dynamism in military facilities because they can temporarily disrupt human and financial resources within a facility. Government shutdowns impacting federal employee pay can sometimes require select civilian employees to be furloughed or work without pay (Burwell, 2013). These furloughs can negatively impact staffing levels, strain operations, and make facilities more vulnerable to poor outcomes. For example, low nurse staffing levels have been associated with nurse burnout (Spence Laschinger & Leiter, 2006) and poorer patient outcomes (Lang, Hodge, Olson, Romano, & Kravitz, 2004; Spence Laschinger & Leiter, 2006). Shutdowns may also force organizations to prioritize the essential tasks of care delivery over performance improvement efforts. Even when government shutdowns are short-lived or are threatened but do not materialize, they can cause disruptions in resources. Organizations may conserve or curtail nonessential resource expenditures in order to manage uncertainty about a future government shutdown. These resource disruptions can potentially impact hospital performance, particularly if resources are shunted from performance improvement activities to cover essential healthcare services during periods of shutdown or threatened shutdown.

Conceptual Model

RDT predicts that when MTF funding is dependent upon quality performance, facilities will have a motive to improve performance in order to access more financial resources for that facility. Thus, facilities will enact performance improvement efforts to improve performance and manage access to critical financial resources. The change in quality performance is also impacted by characteristics of the resource environment, including munificence, dynamism, and complexity. (see Figure 5 for a depiction of the RDT framework that is applicable to this study).

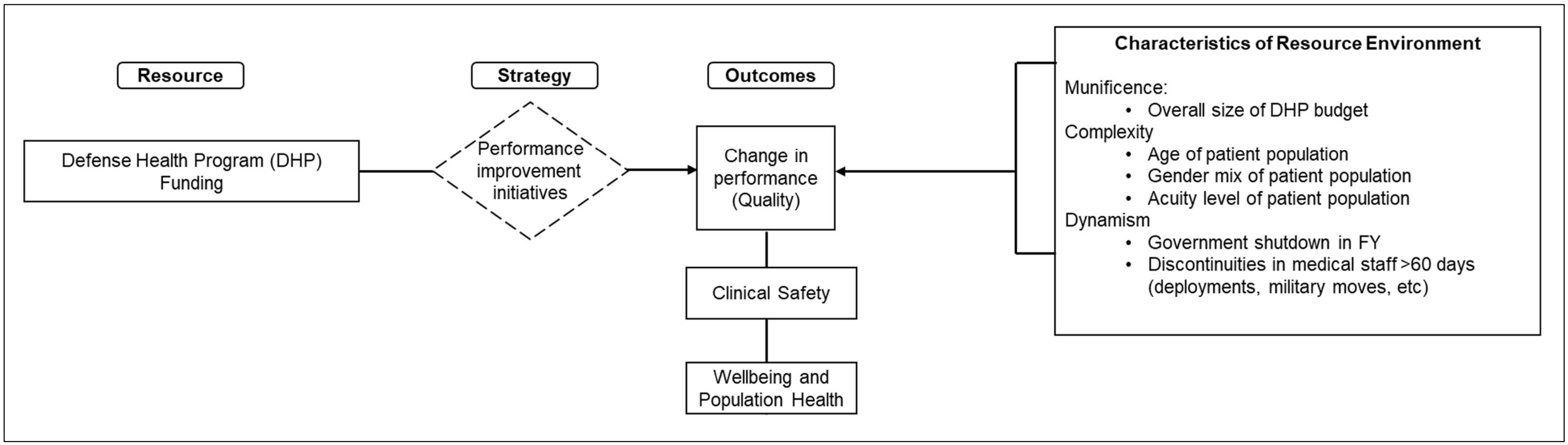

Figure 5. Conceptual framework using RDT as its basis.

Hypotheses

When financial incentives are contingent upon performance, it creates a motive for organizations to focus attention on performance improvement activities (Ocasio, 1997; Pfeffer & Salancik, 1978). Through financial incentives, the entity responsible for establishing the performance criteria and disbursing the financial rewards is able to influence the practices and priorities of healthcare organizations (Ramanujan & Rousseau, 2006). In the private sector, third-party payers motivate hospitals based on the revenue they provide to those organizations. When payers institute pay for performance programs, it creates a motive for healthcare providers to improve performance on the measures incentivized by those payers. Similarly, within the MHS, the medical departments of the Army, Navy, and Air Force exert influence over MTFs in their commands through the resources they provide to those facilities. When MEDCOM instituted its PBB programs (PBAM and IRIS), it sought to motivate Army MTFs to improve performance on selected quality measures in order to obtain greater funding levels. Because Army hospitals are dependent on DHP funds, RDT predicts that Army facilities will adopt performance improvement strategies if funding is contingent upon quality performance.

In facilities with no PBB programs (e.g., Navy and Air Force MTFs), there is less incentive to devote resources to quality improvement because funding levels are not tied to performance. Thus, the following is hypothesized:

H1: There will be greater performance improvement on incentivized quality metrics in Army healthcare organizations that participated in PBB programs relative to military healthcare organizations in other branches that did not participate in PBB programs, ceteris paribus.

The effects of the PBB programs (PBAM and IRIS) may not be limited to the immediate postincentive period. Though the greatest performance change is expected to occur closest to the time in which each specific performance measure became incentivized, the effects may last well beyond that time period. As long as the incentives remains in place, RDT predicts that PBB programs will maintain ongoing pressure for facilities to sustain performance. Facilities are expected to maintain performance in order to obtain the highest possible funding levels. Thus, the following is hypothesized:

H2: Quality improvements attributable to PBB programs in PBB facilities will be sustained throughout the post period after each metric is incentivized.

Summary

RDT predicts that organizations will adopt strategies to help manage access to critical resources in the environment. This chapter established a theoretical framework by drawing on RDT to predict the impact of PBB on quality performance in U.S. Army healthcare facilities. Two testable hypotheses were proposed based on the theoretical framework. The next chapter outlines this study's methods and approach, including the research design, study sample, data sources, operationalization of variables, analytic methods, and sensitivity analysis.

Chapter 4: Methodology

This chapter outlines the research methods used to explore the effects of PBB on quality metrics in military healthcare facilities. The first section describes the research design and the rationale behind its use. The next two sections identify the data sources and study sample, followed by a description of how all of the variables were measured and constructed from the available data. The next two sections discuss the empirical methodology and analytic strategies used to examine the study's research questions. The final section describes the sensitivity analyses used to assess the robustness of the empirical results.

Research Design

This study is a longitudinal post-hoc analysis of quality performance in military healthcare facilities using a quality performance dataset that spans a 15-year time frame. This study aims to assess the impact of PBB on quality performance (Research Question 1) and determine how quality performance is sustained over time after the initiation of PBB (Research Question 2).

To test the hypotheses outlined in the previous chapter, this study uses an organization-level (MTF) difference-in-differences (DID) analysis. The DID approach is used because it enables a quasi-experimental post-hoc analysis of the U.S. Army's two PBB programs. The credibility of the DID design has been widely established in a variety of studies involving health policy analysis (Angrist & Pischke, 2010; Ryan, Burgess, & Dimick, 2015). It is useful in this context because it represents the best approach for estimating the potential effect of PBB on quality performance when examining observational data by comparing facilities exposed to PBB programs to facilities not exposed to PBB programs before and after program implementation.

The study encompasses two separate analyses for Research Questions 1 and 2, each using a slightly different estimation approach. Research Question 1 (main analysis) examines whether the Army's PBB programs are linked to a change in quality performance, ceteris paribus. This research design employs an approach similar to prior research by R. M. Werner, Konetzka, and Polsky (2013) using a DID model to compare Army treatment facilities (subject to PBB incentives) to a comparison group comprised of Navy and Air Force treatment facilities (that did not employ PBB programs). For Research Question 1, quality performance is compared for both groups in the 1- to 3-year period immediately prior to the implementation of PBB incentives. Quality performance is also compared between both groups up to 5 years after the implementation of the PBB programs. Since each quality measure became incentivized through the Army's PBB programs at different points in time, the pre- and postimplementation periods will differ from measure to measure based on when PBB incentives were first offered for each measure. For some performance measures, 5 years of postincentive performance data are not yet available. For these measures, a shorter period of post-PBB performance data is included in the analysis.

In order to assess how performance changed over time in postimplementation period (Research Question 2), performance in both the treatment and comparison groups is also analyzed for two different postimplementation time periods for those quality metrics with 5 years of postimplementation data. The first postimplementation period encompasses the first 10 quarters (2.5 years) after the PBB incentives were initially offered to the treatment group.

Quality performance is also assessed in the second postimplementation period for quarters 11-20 (2.5-5 years) after PBB incentives were offered to the treatment group.

Data Sources

Data for this study are retrieved and merged from a variety of sources to develop a dataset that encompasses a wide range of administrative, demographic, and performance information about each healthcare facility. The two primary data sources are the MDR and the Carepoint Information Portal. The MDR is a centralized data repository that receives, archives, and validates DHA corporate healthcare data from military healthcare facilities worldwide (DHA, 2019f). It uses standardized data processing methods approved through the DHA (2019f) to ensure that health data are collected and managed in a consistent manner across all DOD healthcare organizations. The MDR captures individual, patient-level data about all healthcare encounters that occur within the MHS.

For the current study, de-identified patient-level data was aggregated for each MTF to obtain facility-level data on patient demographics, patient acuity, and provider continuity. Individual-level data were aggregated using statistical software available on a secure virtual server linked to the MDR through the Army's Person-Data Environment, which is an Army and DOD secure business intelligence platform that enables researchers to access unclassified information on active duty, reserve, and retired servicemembers and their dependents using an extensive transformation process that minimizes the risk of identifying individuals (Vie, Griffith, Scheier, Lester, & Seligman, 2013; Vie et al., 2015). Specific information about the construction of these variables is provided in a later section.

Facility-level performance data are obtained from the Carepoint Information Portal. Carepoint is also operated by the DHA and contains facility-level performance data for a wide range of quality performance metrics (DHA, 2019b). It is the primary platform used by clinical leaders at MHS facilities to monitor and track performance trends for improvement initiatives and programs such as the Army's PBAM and IRIS programs. The measures reported through the Carepoint Information Portal are metrics commonly used in healthcare performance evaluation in the private sector. Examples include the HEDIS (NCQA, 2019), ORYX National Hospital Quality Measures (The Joint Commission, 2018), the National Perinatal Information Center (2019) outcome measures, and the National Surgical Quality Improvement Program (American College of Surgeons, 2019). Data in the Carepoint Information Portal are validated through partnerships between the DHA and the healthcare quality agencies that develop the measures and manage the data. For example, the MHS HEDIS data are managed and validated through the NCQA and disseminated to clinical leaders at each facility through the Carepoint portal.

Administrative data about the healthcare facilities, such as location, state, market regions and bed sizes, are retrieved from the health.mil website. This website is updated monthly by the DHA and is publicly available (DHA, 2020). Data are merged across all sources using the unique Defense Medical Information System identifier (DMIS ID). The DMIS ID numbers are four-digit codes attached to all military healthcare facilities. These codes are the controlling standard for facility identification in cost, manpower, and healthcare applications across the DOD (DHA, 2020), which makes it possible to accurately merge data across multiple information systems.

Study Sample

Data for this study are drawn from direct care (i.e., military operated) healthcare facilities in the MHS. As of 2019, the MHS is comprised of a total of 475 health facilities

(DHA, 2019e), but 47 are excluded. Exclusion criteria are established to ensure the homogeneity of the sample and reduce the likelihood of confounding influence from variables unrelated to the research questions. Coast Guard clinics and facilities jointly operated by the DOD and VHA are excluded from analysis due to significant differences in management structure and funding, which would have made them an inappropriate for comparison to hospitals affected by PBB. Navy Operational Force clinics and military Aid Stations are also excluded from the sample due to the fact that these clinics are under minimal management by the MHS and unlikely to be impacted by funding changes or policies enacted through the MHS. Behavioral health clinics and occupational health clinics were excluded because they provide a limited scope of services that were largely unrelated to the performance measures examined in this study.

The majority of the 428 facilities in the study sample are located on or around military installations in the United States ($n = 334$) and abroad. The facilities located outside of the United States are spread across various countries in Europe ($n = 53$), Latin America ($n = 3$), and the Pacific region ($n = 38$).

Sample healthcare facilities are collectively referred to as Military Treatment Facilities or MTFs but are divided into three types: health clinics, hospitals, and medical centers. Health clinics are defined as medical facilities that do not provide inpatient services. Health clinics typically offer primary care and a limited number of specialty care services (DHA, 2019g). Hospitals are defined as facilities that offer both inpatient and outpatient services and a variety of specialty care services. Medical centers are defined as large hospitals that offer both inpatient and outpatient care, as well as a combination of specialty and subspecialty care services. Medical centers typically operate Level 1 Trauma Centers that are authorized to provide emergency care for non-military patients in their respective communities. Due to the diversity of specialty care provided, medical centers also usually host GME programs and medical research (DHA, 2019g).

The majority of the sample facilities are clinics (*n* = 377), many of which are subordinate to larger healthcare organizations, or *parent facilities*. The remaining sample facilities offer inpatient services (*n* = 51) and are comprised of a combination of hospitals (*n* = 34) and medical centers (*n* = 17). As of 2020, the number of inpatient beds in sample medical centers ranges from 47 to 382 with an average of 155. The number of beds in sample hospitals range from 12 to 132 with an average of 59.

All facilities in the sample are principally operated by one of the three major service components (i.e., Army, Navy, or Air Force). As of 2018, all facilities are aligning under the operational control of a single entity, the DHA, although they are still primarily staffed and operated through their respective military branches. Within this study sample, there are 88 Air Force facilities, 187 Army facilities, and 153 Navy facilities. There are also five facilities in the sample that are jointly operated by two or more services (referred to as *joint* facilities) and 132 facilities operating in MSMs.

The sample is subdivided into two groups based on inclusion in PBB programs. The intervention group (*n* = 187) is comprised of Army facilities that participated in PBB programs during the study period (2006-2018). The comparison group (*n* = 241) is comprised of Navy, Air Force, and combined-service facilities that were not included in PBB programs during the study period.

The choice of these particular comparison groups helps to address the potential confounding effects commonly arising from post-hoc secondary analyses that do not employ a randomized control design (Ryan, Burgess, & Dimick, 2015). The intervention group is comprised of Army facilities that were exposed to PBB. Navy and Air Force hospitals were chosen for the comparison group because they are part of the same health system, they serve similar patient populations and they are financed through the same funding source (the DHP budget). Facilities in both groups also support U.S. military missions, which means they face similar organizational stressors that impact their resource environment (such as rapid deployment of military staff and government shutdowns). Facilities in both groups also have similar regulatory environments, since many of the governing regulations and policies are implemented at the DOD level and apply to all three services. Finally, all facilities share a common performance management platform, the Carepoint Information Portal (DHA, 2019b). These similarities help to reduce validity threats that may arise due to pre-intervention differences between the groups that are unrelated to PBB. Table 2 provides descriptive information about the facilities in each group.

Table 2

Description of Facilities in Study Sample

Variable			Facilities
Location			
	United States		
		East	208
		West	126
	Outside U.S.		
		Europe	53
		Pacific	38
		Latin America	3
Facility type			
		Medical center	17
		Hospital	34
		Clinic	377
Sample totals			428
Inpatient beds (2020 mean)			
		Medical center	155
		Hospital	59

The demographics of the patient population reflect approximately 3.4 million patients who received care at each facility during the study period and do not necessarily reflect the demographics of the full population of 9.5 million patients eligible for care within the MHS. For example, active duty patients and their dependent family members are overrepresented in military facilities because they are given the highest priority for enrollment in the direct care system, followed by retirees and their family members. Thus, the average age of patients enrolled at each facility in the sample is 28 with demographics skewed toward the18-24 age band (25.1%) and the 25-34 age band (28.3%), which heavily reflects the active duty population and their family members. Additionally, facilities in this study are, on average, comprised of 65% male patients. This is likely due to the fact many clinics exclusively serve active duty servicemembers, and only 16.5% of the active duty U.S. military force are female (DOD, 2018b). Table 3 provides more detailed demographic information for study facilities.

Table 3

Patient Demographics for Study Facilities

		Group	
	Variable	Sample	Population
Gender (%)			
	Male	64.5	51.0
	Female	35.5	49.0
Age (%)			
	< 4	8.1	6.0
	5-14	7.1	11.4
	15-17	1.4	3.3
	18-24	25.1	12.3
	25-34	28.3	12.5
	35-44	17.2	9.2
	45-64	11.0	22.3
	65+	1.7	24.1
Acuity	RVUs per patient per year	130.0	*

Note. RVUs = relative value unit. Population refers to the total population of patients enrolled in the TRICARE program during FY 2019. This includes patients who receive care at civilian (non-DOD) healthcare facilities.

[1] From Evaluation of the TRICARE Program: Fiscal Year 2019 Report to Congress.

It is important to note that the sample size fluctuates for each performance metric based on the number of applicable facilities in existence during the relevant timeframe. This reflects the fact that the number of MTFs in the MHS has fluctuated over the study period in response to the operational needs of the U.S. military, coupled with the implementation of Community-Based Medical Homes (Carabajal, 2012) and the closing/restructuring of inpatient healthcare facilities. The sample size also fluctuates for some performance measures due to the scope of services offered at the facilities and the number of patients eligible for those services. For example, the CLABSI performance measure includes a significantly smaller sample size in comparison to the population-health measures because only 51 of the sample facilities have inpatient capabilities. Additionally, facilities are excluded from the analysis of each performance measure if there are fewer than 5 patients receiving care at that facility that were eligible for the relevant services (e.g., mammographies).

Variable Measurement

The following section provides information about the construction of key variables, including descriptions of dependent variables, model covariates, and the key independent variable. The dependent variables and model covariates are presented first because they are the same for both Research Questions 1 and 2. The key independent variable is presented in the third section because it differs for each research question.

Variable Construction: Dependent Variables

Both research questions investigate whether participation in the Army's PBB programs is associated with a change in quality performance for incentivized metrics. Over 40 quality measures were incentivized at various times in the PBAM and IRIS programs. These were selected for inclusion in the PBB programs based on the Army's strategic priorities and their applicability to the MHS Quadruple Aim. The MHS Quadruple Aim encompasses four domains of which three follow: improving population health, per capita cost, and experience of care (Berwick et al., 2008). The fourth domain of the Quadruple Aim, readiness, is specific to the U.S. military and is defined as "ensuring that the total military force is medically ready to deploy and that the medical force is ready to deliver healthcare anytime, anywhere in support of the full range of military operations, including humanitarian missions" (DHA, 2013).

Though quality, cost, efficiency, productivity, and experience of care measures were included in both the PBAM and IRIS programs, the current study focuses specifically on quality measures. The specific measures under investigation are a subset of the total list of quality measures that were incentivized over the life of the program. The dependent variable (quality improvement) is constructed using 13 of the 40 quality metrics that were incentivized through the Army's PBB programs. Performance measures are selected with incentive start dates that span the timeframe of the PBAM and IRIS programs. This facilitates examination of the programs' effects over time and helps to reduce threats to internal validity caused by historical artifacts. Performance measures are also selected based on the availability of data. Measures with less than four quarters of pre- or postintervention data are excluded. This exclusion primarily pertains to measures that were first incentivized near the start of the PBAM program in 2006 or very recently in the IRIS program. Measures are also excluded if data are not available for both the intervention and comparison groups. This exclusion pertains to certain measures that were specific to Army facilities.

For example, the healthy weight performance metric measures the proportion of patients in each facility with a Body Mass Index value recorded during primary care medical encounters. This metric is excluded from the study because data were collected within Army facilities but not uniformly collected and reported in the comparison facilities. A third exclusion criterion is applied to eliminate instrumentation threats to internal validity. Performance measures are eliminated if they did not use the same measurement instrument for all facilities in both the intervention group and the treatment group. For example, all patient satisfaction measures are eliminated because the service branches did not start using a common satisfaction survey until 2014. Figure 6 presents a diagram of the selection process for performance measures selected for study inclusion.

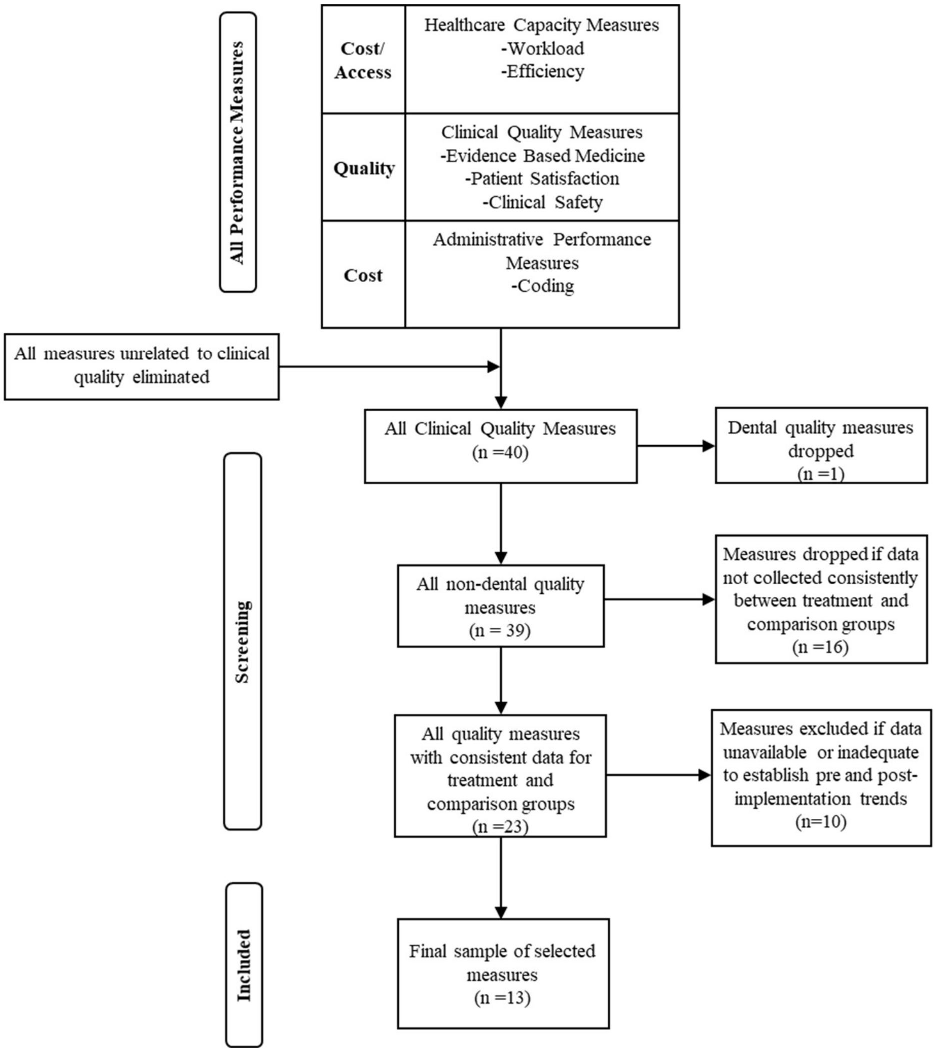

Figure 6. Diagram depicting selection criteria for performance measures.

All performance metrics that comprise the dependent variable are evaluated independently (i.e., a composite score is not used). They are measured at the facility level and are expressed as facility rates or proportions of patients obtaining care at each facility. There are two types of performance metrics: *population health* measures and one *clinical safety* measure.

Population health. The well-being and population health domain pertains to all outpatient facilities that offer primary care services. This variable indicates facility-level performance on measures that assist in the management of chronic conditions and promote well-being and healthy behaviors. The population health domain is comprised of 12 measures derived from the HEDIS. The HEDIS dataset is a group of quality performance measures that are validated, supported, and maintained through a partnership between the DHA and the NCQA. HEDIS measures are used to analyze quality performance in managed care organizations in domains such as managing chronic disease, delivering preventive care, and managing acute illness (U.S. Army Medical Command, 2016).

Population health measures are investigated in this study because they are an important part of promoting health system accountability in improving the overall health of communities. The use of population health measures, particularly those measures pertaining to preventive health, are intended to reduce costs by keeping patients healthier, reducing hospitalizations, and reducing unnecessary utilization (Stoto, 2017). HEDIS measures are chosen in particular because these measures are some of the most widely adopted healthcare performance measurement tools in the United States (NCQA, 2019). Studies have indicated that HEDIS population health measures are associated with better health outcomes, particularly among populations of patients with chronic diseases such as diabetes (Harman et al., 2010). The use of HEDIS measures can also promote cost-effective care by reducing future costs associated with serious illness, although the measurement of such benefits varies widely (Neumann & Levine, 2002). Additionally, prior evidence suggests that pay for performance has had significant effects in diabetes care and disease management in primary care settings (Gupta & Ayles, 2019; Lin et al., 2016), though no studies are yet available to determine effects in conjunction with PBB.

Three of the selected HEDIS population health measures specifically pertain to children. Though these measures do not directly address healthcare quality pertaining to military servicemembers, they do reflect the comprehensive nature of the overall MHS mission. As discussed previously, one of the two primary missions of the MHS is its purpose in providing medical benefits to retirees and dependent family members of military servicemembers. The three measures pertaining to children are selected because they help to assess each facility's performance in providing care that is specifically tied to the dependent family members as opposed to the active force.

Clinical safety. The clinical safety domain pertains to all inpatient facilities in the study sample. It is constructed using the rate of CLABSI, which is one type of HAI commonly addressed in the quality literature. Although some researchers have noted inconsistent results in pay for performance related to HAIs (Vokes et al., 2018), several studies have demonstrated positive effects (Bastian et al., 2016; Calikoglu et al., 2012). This measure is selected for the clinical safety domain in order to add to the body of research on pay for performance and HAIs.

The CLABSI measure is defined as the risk-adjusted rate of CLABSIs at each facility, relative to predicted rates of CLABSI events based on data collected from hospitals participating in the NHSN program sponsored by the CDC (2019; U.S. Army Medical Command, 2017). A central line is defined by the CDC as "an intravascular catheter that terminates at or close to the heart or in one of the great vessels which is used for infusion, withdrawal of blood, or hemodynamic monitoring" (Powell, 2018). Further, a central line that is eligible for inclusion in the CLABSI measure is defined as

> a central line that has been in place for greater than two consecutive calendar days following the first access of the central line, in an inpatient location, during the current admission. Central lines are eligible for CLABSI events until the day after removal from the body or patient discharge, whichever comes first. (Powell, 2018)

The MHS collects and reports its CLABSI data using the protocols developed and validated by the CDC. The CLABSI measure incentivized through the Army's IRIS program corresponds to the CDC's CLABSI standardized infection ratio (SIR), which "is calculated by dividing the number of observed events by the number of predicted events. The number of predicted events is calculated using probabilities estimated from negative binomial models constructed from 2015 NHSN data, which represents the baseline population" (CDC, 2020, p. 4-28). Table 4 provides detailed information about the construction of the dependent variable and the relevant time frames.

Table 4

Dependent Variable Construction and Data Collection Periods

Measure	Incentive start	Pre-	Post-	Variable type	Definition
Aim 1: Population Health					
Mammography	2006	2004 Q4-2006 Q3	2006 Q4-2011 Q4	Population Health	% eligible patients seen at the facility with mammography within the last 24 months.
Cervical cancer screening	2006	2004 Q4-2006 Q3	2006 Q4-2011 Q4	Population Health	% eligible patients seen at each facility with cervical cancer screening within the past 36 months.
Diabetes-A1C screening	2006	2004 Q4-2006 Q3	2006 Q4-2011 Q4	Population Health	% eligible patients seen at each facility, diagnosed with diabetes, who have had A1C testing within the past 12 months.
Diabetes-A1C control	2006	2004 Q4-2006 Q3	2006 Q4-2011 Q4	Population Health	% eligible patients seen at the facility, diagnosed with diabetes, with A1C levels at nine or below.
Diabetes-LDL control	2006	2004 Q4-2006 Q3	2006 Q4-2011 Q4	Population Health	% eligible patients seen at each facility, diagnosed with diabetes, who have an LDL level below 100.
Asthma care	2006	2004 Q4-2006 Q3	2006 Q4-2011 Q4	Population Health	% patients with persistent asthma who are prescribed medications considered acceptable as a primary therapy for the long-term control of asthma.
Well child visits (6 in 15 mos.)	2012	2010 Q1-2011 Q3	2011 Q4-2016 Q4	Population Health	HEDIS criteria for percent of eligible population with 6 or more well child visits in the first 15 months of life.
Low back pain	2016	2013 Q4-2015 Q3	2015 Q4-2018 Q4	Population Health	% eligible patients seen at each facility, aged 18-50, with a primary diagnosis of low back pain who did not have an imaging study (plan x-ray, MRI, CT scan) within 28 days of diagnosis.
Appropriate testing for children with pharyngitis	2017	2015 Q1-2016 Q3	2016 Q4-2018 Q4	Population Health	% children 2=18 years of age who were diagnosed with pharyngitis, dispensed an antibiotic and received a group A streptococcus test for the episode. A higher rate represents better performance (i.e., appropriate testing).

Measure	Incentive start	Pre-	Post-	Variable type	Definition
Appropriate treatment for children with URI	2017	2015 Q1-2016 Q3	2016 Q4-2018 Q4	Population Health	% children 3 months to 18 years of age who were given a diagnosis of URI and were not dispensed an antibiotic prescription. A higher rate indicates better performance.
Follow-up after 7 days of hospitalization for mental illness	2017	2014 Q4-2016 Q3	2016 Q4-2018 Q4	Population Health	% adult and children 6 years of age and older who were hospitalized for treatment of selected mental health disorders and received follow-up within 7 days of discharge.
Follow-up after 30 days of hospitalization for mental illness	2017	2014 Q4-2016 Q3	2016 Q4-2018 Q4	Population Health	% adult and children 6 years of age and older who were hospitalized for treatment of selected mental health disorders and received follow-up within 30 days of discharge.
Aim 1: Clinical Safety					
CLABSI	2017	2015 Q2-2017 Q3	2017 Q4-2018 Q4	Clinical Safety	Risk-adjusted rate ratio of CLABIs in MTF ICUs compared to other participating ICUs in the CDC NHSN Program.
Aim 2					
Mammography	2006	2003 Q4-2006 Q3	2006 Q4-2011 Q4	Population Health	% eligible patients seen at the facility with mammography within the last 24 months.
Cervical cancer screening	2006	2003 Q4-2006 Q3	2006 Q4-2011 Q4	Population Health	% eligible patients seen at each facility with cervical cancer screening within the past 36 months.
Diabetes-A1C screening	2006	2003 Q4-2006 Q3	2006 Q4-2011 Q4	Population Health	% eligible patients seen at each facility, diagnosed with diabetes, who have had A1C testing within the past 12 months.
Diabetes-A1C control	2006	2003 Q4-2006 Q3	2006 Q4-2011 Q4	Population Health	% eligible patients seen at the facility, diagnosed with diabetes, with A1C levels at nine or below.
Diabetes-LDL control	2006	2003 Q4-2006 Q3	2006 Q4-2011 Q4	Population Health	% eligible patients seen at each facility, diagnosed with diabetes, who have an LDL level below 100.
Asthma care	2006	2003 Q4-2006 Q3	2006 Q4-2011 Q4	Population Health	% patients with persistent asthma who are prescribed medications considered acceptable as a primary therapy for the long-term control of asthma.

Variable Construction: Model Covariates

As discussed previously, there are several factors related to munificence, dynamism, and complexity that can impact a facility's performance and response to PBB incentives. Measures for each of these factors are operationalized as follows, and are included as facility-level covariates in multivariate regression analyses.

Complexity. The covariates associated with complexity are age, gender, and the acuity level of the patients seen at each facility. These variables are based on patient-level variables identified as risk adjusters in previous research (Iezzoni, 2003). Age and gender mix data for each facility are constructed from patient-level data obtained from the MDR for each study quarter. The individual-level data are aggregated to the facility level based on the medical encounters that occurred at each facility during the specified quarter. The gender variable is expressed as the proportion of patients seen at each facility who are female. The age variable is expressed as the average age of patients seen at each facility during each study quarter.

The overall health status of patients, measured in various ways, is widely accepted as a risk adjuster in health services research (Elixhauser, Steiner, & Harris, 1998; Iezzoni, 2003; Kane & Radosevich, 2011) due to the impact that it can have on individual health outcomes (Kane & Radosevich, 2011). For the current study, a proxy variable for the acuity level of patients seen at each facility is constructed using the total RVUs added up for a patient's care across all medical facilities over a 12-month period. RVUs provide a standardized measure of the resources that are used to provide patient services (National Health Policy Forum, 2015). An assumption of this study is that the complexity and intensity of medical care required for each patient is roughly proportional to the RVU value of the patient's medical encounters.

Thus, the total sum of a patient's RVUs over a specified period of time can provide an approximation of the patient's medical acuity.

The acuity variable is constructed by adding each patient's aggregate RVUs (work and practice expense) for a full calendar year. In order to avoid endogeneity concerns, the acuity variable is then lagged by 1 year for each patient, so that the patient's acuity represents the total RVU's for the year prior. Specifically, each patient's individual acuity is expressed as the sum total of all RVUs generated by that patient across all medical facilities in the year prior. The facility-aggregated acuity variable represents each facility's average acuity for all patients seen in a particular quarter.

Dynamism. Covariates related to dynamism are government shutdowns and provider discontinuities. Government shutdowns are indicated with a dummy variable equaling 1 during periods with an active government shutdown and 0 otherwise. Data regarding periods of government shutdowns are collected from publicly available historical records.

The variable for provider discontinuities is constructed using a methodology similar to previous research on provider continuity in U.S. military medical facilities (Schwab, 2018). Staff discontinuity is operationally defined for this study as any period of 60 days or greater in which a medical provider (staff member with a National Provider Identifier) records zero medical encounters. This variable is constructed using encounter data from the MDR. The first step is to count the number of medical encounters recorded for each National Physician Identifier code at each facility for each month. Physician discontinuities are identified if a provider records zero patient encounters in a particular month after having recorded at least one encounter in the same facility during the preceding month (i.e., discontinuity = 1 for each month a provider records zero encounters). Discontinuities are recorded for the first 60 consecutive days in which a provider records zero patient encounters in a facility in which he or she had previously seen patients. To avoid biasing the discontinuity estimates, a provider is no longer counted as a discontinuity after 60 consecutive days of recording zero medical encounters. The rationale for this exclusion is that if the provider has no medical encounters after the 60-day window, this is an indication that he or she no longer works at a particular facility rather than indicating a disruption in providing care for a short period of time.

Provider discontinuities are aggregated to the facility level by calculating the ratio of provider discontinuities divided by the total number of active providers in each facility in each time period. Providers are considered to be active in a facility if the National Physician Identifier code is associated with at least one medical encounter occurring in that facility. As mentioned, providers are dropped from the facility's count of active providers after 60 consecutive days of recording zero medical encounters in that facility.

Munificence. Munificence is operationalized according to the yearly inflation-adjusted DHP budget for all facilities. Data for the DHP budget is obtained from the public archives of the NDAA for each FY (DOD, 2019). The DHP budget is then adjusted for annual inflation using the general Consumer Price Index for each period and transformed to a logarithmic value. Table 5 provides detailed information regarding the construction of model covariates and variables of interest, along with data sources for each variable.

Table 5

Overview of Key Variables and Model Covariates

Variable	Data source	Construction
Variables of Interest (IV)		
PBB		Coded as 1 if the facility is subject to PBB incentives through the PBAM or IRIS programs, 0 otherwise.
Aim 1		
Post		Coded as 1 if the observation occurs during the postintervention period, 0 otherwise.
Aim 2		
First		Coded as 1 if the observation occurs in quarters 1-10 of the postintervention period.
Second		Coded as 1 if the observation occurs in quarters 11-20 of the postintervention period.
Facility-level covariates		
Complexity		
Age (Average)	MDR (2004-2018)	Average patient age for all patients seen at facility *j* at time *t*.
Age (Band)	MDR (2004-2018)	Proportion of patients seen at facility *j* at time *t* in each age band:
Band 1 (Reference)		0-4 years
Band 2		5-14 years
Band 3		15-17 years
Band 4		18-24 years
Band 5		25-34 years
Band 6		35-44 years
Band 7		45-64 years
Band 8		65+ years
Gender	MDR (2004-2018)	
%Female		Percent of patients who are female.
Acuity (00s)	MDR (2004-2018)	Average lagged total RVU's for each patient seen at facility *j* at time *t-1,* in 100s.
Dynamism		
Shutdown	Public Records (2004-2018)	Coded as 1 if a government shutdown occurred during the time period, 0 otherwise.
Provider Discontinuity	MDR (2004-2018)	Percent of active medical providers who recorded 0 patient encounters for a period >60 days.
Munificence		
DHP budget	NDAA (2004-2018)	Natural log of the inflation-adjusted DHP budget for each year.

Variable Construction: Key Independent Variable

The key phenomenon under investigation in this study is the exposure to PBB incentives through the Army's PBAM and IRIS programs. Exposure to PBB incentives is indicated by a dichotomous variable (1 = PBB) for Army facilities involved with the program. For Research Question 1, the key independent variable is an interaction between PBB = 1 (for Army facilities participating in PBB) and Post = 1, which represents the years for which the PBB incentives were active for each measure. The pre- and postintervention periods vary for each outcome measure based on the timeframe in which the measure first became incentivized and data availability. Table 4 provides specific information about the pre- and postintervention periods for each measure and each research question. For Research Question 1, the postintervention period is defined as up to 5 years after the measure was incentivized.

For Research Question 2, the postintervention period is divided into two separate periods for the seven performance metrics in Table 4 that have 5 years of postimplementation data. The first postintervention period is indicated by a dichotomous variable (1 = First) if the observation occurs during the first 2.5 years (10 quarters) after the measure became incentivized. The second postintervention period is indicated by a different dichotomous variable (1 = Second) if the observation occurs 2.5-5 years (quarters 11-20) after the initiation of the incentives. The key independent variable for Research Question 1 is an interaction between PBB = 1 (for Army facilities participating in PBB) and First = 1 (first 10 quarters postincentive), and PBB = 1 and Second = 1 (second 10 quarters postincentive).

The main analysis for Research Question 1 estimates the marginal effect of the PBB incentives by comparing pre-/post-differences in the performance for facilities exposed to the program (Army facilities) to those facilities in the comparison group that were not exposed to

PBB incentives (Air Force and Navy facilities). The main analysis for Research Question 2 estimates the marginal effect of PBB incentives during two distinct time periods after the initiation of the performance incentives (the first 10 quarters and the second 10 quarters). As indicated previously, the purpose is to determine how performance changes over time in the postintervention period.

Empirical Methodology

The research design for both the main analysis (Research Question 1) and the secondary analysis (Research Question 2) is an organization-level (MTF) DID analysis. The DID approach is used because it enables a quasi-experimental analysis of post-hoc data that were collected to support the U.S. Army's two PBB programs. This approach addresses the potential validity concerns associated with post-hoc analysis by assessing performance trends before and after program implementation using similar comparison groups.

Inferences about the effects of PBB are drawn using two key assumptions of the DID analysis. First, the parallel trends assumption requires that the pattern of performance of both groups must be similar prior to the program implementation. The assumption is that, in the absence of treatment or intervention, any differences in performance between the two groups would hold constant over time (Angrist & Pischke, 2008). To determine if this assumption is reasonable, performance tends for the intervention and comparison groups are plotted over time and visually inspected for parallel trends.

The second critical assumption of the DID is the stable unit treatment value assumption (Rubin, 1977), which requires that there is only one form of the treatment (exposure to PBB or no exposure to PBB) and that there is no interference between the groups (spillover effects). The potential for spillover effects is assessed using a sensitivity analysis in which treatment and comparison facilities operating in close proximity to one another (in MSMs) will be eliminated from the analysis.

The study uses a Poisson panel regression to conduct the analyses for Research Questions 1 and 2. The nature of the data is most appropriate for a Poisson or negative binomial regression model because the dependent variable is comprised of a discreet count of the number of patients eligible for a particular type of treatment or service and a discreet count of the number of patients that received the prescribed care. For the CLABSI variable, a negative binomial regression may have been more appropriate given the large number of observations that were zero (patients did not get an infection). However, the model did not converge during preliminary analyses, indicating that the data were a poor fit for this model.

Research Question 1 tests whether the Army's PBB programs are linked to a change in quality performance, ceteris paribus, using a research design similar to prior research by R. M. Werner et al. (2013). In this case, the DID model is applied to compare Army treatment facilities (subject to PBB incentives) to a comparison group comprised of Navy and Air Force treatment facilities (that did not employ PBB programs). The analysis for Research Question 2 is an extension of Research Question 1, examining two distinct postperiods for those indicators with sufficient postintervention data. This second analysis addresses whether performance improves, declines, or remains the same throughout the 5-year postintervention period, ceteris paribus.

Preliminary Analyses

Several preliminary analyses are performed. First, the data are analyzed carefully to identify any missing data. Large sections of missing data are investigated to determine if there are indications of merge errors. A variable for merge errors is retained in the master dataset for each merge that is performed so missing data occurring through merge errors can be easily identified and addressed. In most cases, missing data are attributable to clinics that either activated or deactivated during the study period. This is identified through an "IA" code in the facility name. Next, the data are carefully analyzed to identify and address extreme or inappropriate values. Box plots, distribution tables, and other methods are used to examine the spread of data, which are subsequently used to identify extreme values and outliers. Extreme values are assessed individually and excluded from the study sample when appropriate. For example, extreme values in the performance variable usually occur in clinics with very few patients eligible for the relevant care. In most cases, this is due to the fact that the clinic was activating or deactivating during the specified time period. Due to this concern, clinics with fewer than 5 patients eligible for care (pertaining to the quality measure) are dropped from the analysis to avoid biasing the results.

Since the majority of dependent variable measures are proportions (i.e., number of patients receiving the appropriate care divided by the total number of patients eligible for the care), data are also examined to determine if all values fall between the bounds of 0 and 1. The ranges of these variables are analyzed to determine the most appropriate regression model. For example, when the values in the dependent variable occur close to the bounds (less than .2 or greater than .8), then a Poisson regression provides more valid statistical tests of coefficients (Long, 1997).

Estimation Approaches

Aim 1

A facility-level, fixed-effect, Poisson regression model is used for Aim 1, except in the cases indicated below. Standard errors are adjusted for clustering at the facility level to account for the fact that observations are repeated for each facility over time and nonindependent (Huber, 1967; Ryan, Burgess, & Dimick, 2015; R. M. Werner et al., 2013; White, 1980).

The equations for study Aim 1 follow:

[Equation 1] $Y_{jt} = b_0 + b_1 Post_t + b_2(PBB_j * Post_t) + b_3 \mathrm{X}_{jt} + u_{j+} \varepsilon_{it}$

All variables are indexed *j* for a given facility and *t* for each quarter. The PBB_j indicator variable is coded as a 1 if the facility was subject to PBB incentives during the observation period and 0 otherwise. The $Post_t$ indicator variable is coded as a 1 if the observation occurred in the years after the implementation of incentives for the specific quality metric under investigation. X_{jt} corresponds to a vector of time-variant, facility-level covariates, including complexity (age, gender, and acuity composition of patients at each facility), facility-aggregated dynamism indicators (percent staff discontinuity, government shutdown (0/1)), and munificence indicators (size of DHP budget). Finally, u_j represents other time-invariant facility characteristics (fixed effects) and ε_{it} is the random error term.

For all Aim 1 analyses, if b_2 is significant and positive, it is concluded that PBB incentives have a positive effect on quality improvement for the given quality measure. As an exception, for areas of performance where a reduction is desirable, it is concluded that PBB incentives have a positive effect on quality improvement if b_2 is significant and *negative.*

[Equation 1a] $Well_{jt} = b_0 + b_1 Post_t + b_2(PBB_j * Post_t) + b_3 \mathrm{X}_{jt} + u_{j+} \varepsilon_{it}$

Equation 1a employs a Poisson regression and includes all HEDIS measures. It applies to all outpatient facilities. $Well_{jt}$ indicates a cluster of performance metrics related to well-being and population health indicators among the patients receiving care at each facility. For each of these measures, dependent variable is generically defined as a count of the number of patients at each facility receiving the appropriate care. The exposure variable is defined as the count of total patients at each facility *j* who are eligible for the indicated care. For example, the dependent variable in the HEDIS mammography measure is the total count of patients at facility *j* who have had mammography services in the past 24 months. The exposure variable (denominator) is the total number of patients at facility *j* who are eligible for mammography services.

[Equation 1b] $ClinSafe_{jt} = b_0 + b_1 Post_t + b_2(PBB_j * Post_t) + b_3 X_{jt} + u_{j+}\varepsilon_{it}$

Equation 1b applies to all inpatient facilities and employs a Poisson regression. The dependent variable is clinical safety and is measured using the count of the observed CLABSI at each facility. The exposure variable is the number of predicted CLABSI events. Predicted CLABSI events are modeled based on national data compiled by the CDC among approximately 25,000 hospitals participating in the NHSN Program (CDC, 2019).

Aim 2

A facility-level, fixed-effect, Poisson regression model is used for all measures in Aim 2. Equation 2 mirrors the analysis of Equation 1, except the $Post_t$ variable is broken into two increments: $First_t$ for first half of the 5-year postperiod (quarters 1-10 after the incentive) and $Second_t$ for second half of the 5-year postperiod (quarters 11-20 after the incentive). $First_t$ and $Second_t$ are be interacted with the PBB_j indicator variable to denote Army facilities targeted by PBB incentives in the first or second 10 quarters of the post implementation period. Equation 2 is used to estimate the effects of the PBB incentives in two distinct periods after the introduction of each incentive in order to determine if the response to incentives was confined to a particular time period (e.g., immediately after the incentive started) or alternatively, if there was a change in the response in the first and second periods after the implementation of PBB. Equation 2 is only applied to the metrics for which 5 years of postincentive data are available.

[Equation 2] $Y_{jt_{1-20}} = b_0 + b_1 First_t + b_2 Second_t + b_3(PBB_j * First_t) + b_4(PBB_j * Second_t) + b_5 X_{jt} + u_{j+}\varepsilon_{it}$

All variables are indexed j for a given hospital and t for each quarter after the incentives were initiated for the given measure. The PBB_j indicator variable is coded as a 1 if the facility was targeted by PBB incentives, and 0 otherwise. The $First_t$ indicator variable is coded as a 1 if the observation occurs in quarters 10-20 after the implementation of incentives for the specific quality metric under investigation. The $Second_t$ indicator variable is coded as a 1 if the observation occurs in quarters 11-20 after the implementation of incentives for the specific quality metric under investigation. X_{jt} corresponds to a vector of facility-level covariates, including complexity (age, gender, and acuity composition of patients at each facility), facility-aggregated dynamism indicators (percent provider discontinuity, government shutdown (0/1)), and munificence indicators (size of DHP budget). Finally, u_j represents other time-invariant facility characteristics (fixed effects) and ε_{it} is the random error term.

For all Aim 2 analyses, statistical tests are conducted to assess the relative magnitudes of estimated coefficients b_3 and b_4. Where an *increase* in the value of a quality metric is desirable, if b_4 is significantly greater than b_3, it is concluded that performance attributable to PBB improved over the course of the postintervention period. Alternatively, if the two coefficients are not significantly different, it is concluded that performance remained the same over the two periods. Finally, if b_4 is significantly less than b_3, it is concluded that initial improvement in the metric diminished. Similar assessments are made for metrics where a decline in value is desirable (i.e., CLABSI) to determine whether performance improved, remained the same, or diminished.

Sensitivity Analyses

Several sensitivity analyses are conducted to determine the robustness of the empirical results with respect to various changes in specification and context that may alter quality performance. The first sensitivity analysis considers the possibility that quality performance may be impacted during an anticipatory period immediately prior to the implementation of incentives. Thus, two sensitivity analyses are performed to determine if facilities alter quality performance in the anticipatory period before incentives are offered. The first sensitivity analysis drops all observations that occur in the 6-month period immediately prior to the effective date of performance incentives. The second sensitivity analysis in this group creates a dichotomous variable indicating observations that occur during the 6-month "anticipatory period" and tests for significance with respect to quality performance.

The second sensitivity analysis considers the possibility that the estimation model is overspecified with the inclusion of the covariates for munificence, dynamism, and complexity. Though the estimates are unbiased, if the model is overspecified, it may have an adverse effect on the variances of the estimators (Wooldridge, 2013, p. 88). The second sensitivity analysis drops all covariates from the model to determine the robustness of the effects of PBB on performance.

A third sensitivity analysis is conducted to examine the impact of operating in a MSM. As discussed previously, MSMs are geographic locations that include healthcare facilities operated by two or more of the uniformed services. Under these MSM conditions, patients are sometimes seen in more than one healthcare facility, often operated by different service branches. This creates the potential for PBB to have spillover effects in local healthcare facilities not directly participating in the program. This is particularly problematic for performance measures involving the management of chronic disease because patient outcomes may be impacted by treatment from a variety of providers in different locations, making it more difficult to isolate the performance effects of the program. To determine if this has an impact on the robustness of the results, a third analysis performs all regressions with the exclusion of facilities operating in MSMs.

As mentioned, observations are dropped for facilities in which fewer than 5 patients are indicated for the care relevant to the performance measure (for example, cervical cancer screenings). In many cases, small numbers are due to the fact that the clinic is activating or deactivating during the specified time period, or the clinic serves a patient population that is largely not applicable to the performance measure (such as a pediatric clinic). A fourth sensitivity analysis examines the robustness of the empirical results using different exclusion parameters for the minimum number of patients eligible for each type of care. Observations are dropped for facilities with fewer than 10 and fewer than 20 patients eligible for care.

The fifth sensitivity analysis provides further examination of the effect that patient age has on the performance results. The age variable used in all regressions in the primary analysis indicates the average age of patients seen at each facility during each specified quarter. Specifying the average patient age is much less precise than using age bands to identify specific age groups that have a larger (or smaller) impact on performance results. It is possible that important age-related effects on performance may be overlooked when patient ages are averaged together rather than specified independently within age bands. Thus, the fifth sensitivity analysis examines a subset of quality metrics for which PBB has demonstrated the anticipated effect of improved performance to assess if replacing the average age variable with variables that indicate patient age bands has an effect on these findings.

Summary

This chapter described the empirical and analytic methods used to evaluate the impact of PBB on quality performance in military healthcare facilities. This study uses a DID approach to analyze post-hoc performance data for two comparable groups of military facilities that either participated in PBB or did not participate in PBB during the specified time period. Data elements from multiple DOD databases are merged with publicly available information to construct the variables used in the facility-level, fixed-effect, Poisson regression models. A variety of sensitivity analyses are also presented to determine the robustness of empirical results.

Chapter 5: Results

This chapter presents study findings based on the methodology described in Chapter 4. It is divided into four sections. The first section provides descriptive statistics regarding the study facilities and patients within those facilities. The following two sections provide the results of the main analysis for Research Question 1 and the sensitivity analyses that were undertaken to test the robustness of the results. The final section describes the results of the main analysis for Research Question 2. A brief summary concludes the chapter.

Results of Descriptive Analysis

As discussed earlier, the study sample consists of a total of 428 healthcare facilities across the DOD. Seventy-eight percent of the study facilities (n = 334) are located within the continental United States, while 22% of facilities are located outside of the United States (n = 94). The majority of study facilities are clinics (n = 377; 88%), followed by hospitals (n = 34; 8%) and medical centers (n = 17; 4%). PBB facilities have slightly higher average bed counts (2020 means) relative to comparison facilities for both community hospitals and medical centers. Table 6 provides an overview of the PBB and comparison facilities with respect to location, facility types, and average number of beds.

Table 6

Descriptive Statistics for Sample Healthcare Facilities (2019-2020)

	Group		
	Comparison	PBB	Total
Location			
United States			
East	111	97	208
West	81	45	126
Outside U.S.			
Europe	23	30	53
Pacific	24	14	38
Latin America	2	1	3
Facility type			
Medical center	8	9	17
Hospital	22	12	34
Clinic	211	166	377
Inpatient beds (2020 mean)			
Medical center	142	169	155
Hospital	54	64	59

Study facilities experienced an average of 10.2% provider discontinuity for each month. There were no significant differences observed between the PBB and comparison groups with respect to provider discontinuity. On average, only 35.5% of patients seen at study facilities were female. As discussed previously, the gender composition reflects the fact that active duty servicemembers are overrepresented in the patient population of sample facilities, and the majority of the U.S. servicemembers are male (DOD, 2018b). The average age for patients seen at study facilities is 28.3 years with no statistically significant differences observed between the PBB and comparison group. The age demographics for both groups are skewed toward the 18-24 and 25-34 age bands, reflecting the overall demographics of the active U.S. military force.

Table 7 compares the means of the PBB and comparison groups across all model covariates. There are statistically significant differences observed between the PBB and comparison groups for some model covariates. For example, there is a statistically significant difference between the two groups across all age bands, but the magnitude of these differences is small. Similarly, the PBB facilities provide care to a slightly higher percentage of women (M = 35.6, SD = 14.0) in comparison to the comparison group (M = 35.4, SD = 11.6), but the magnitude of the difference is also negligible, $t(54390) = -2.06$, $p < .001$. There is also a statistically significant difference between the two groups for acuity, but it is important to note that the standard deviation is very high, indicating a high degree of data spread with respect to this variable.

Table 7

Group Comparison of Means for Model Covariates

	Comparison		PBB			All study facilities	
	M	*SD*	*M*	*SD*	Difference	*M*	*SD*
Complexity							
Gender (%)							
Female	35.4	11.6	35.6	14.0	**	35.5	12.6
Male	64.6	--	64.4	--		64.5	--
Age (%)							
< 4 (reference)	8.2	5.8	7.9	6.6	***	8.1	6.1
5-14	6.9	4.8	7.5	6.2	***	7.1	5.4
15-17	1.3	1.1	1.6	1.8	***	1.4	1.4
18-24	25.4	17.1	24.2	15.0	***	25.0	16.3
25-34	28.5	8.2	28.4	9.0	***	28.5	8.6
35-44	16.8	6.3	17.9	6.8	***	17.2	6.5
45-64	11.2	8.1	10.9	9.0	***	11.0	8.5
65+	1.7	3.0	1.7	3.2	***	1.7	3.1
Age (average)	28.3	4.2	28.3	4.8		28.3	4.4
Acuity							
RVUs per patient per year	149.2	202.5	110.2	123.8	***	134.2	177.524
Dynamism							
Provider Discontinuity (% monthly)	10.2	9.0	10.3	9.2		10.2	9.1

Note. PBB = Performance-based budgeting. SD = Standard Deviation. RVU = Relative Value Unit. Means are aggregated across all study years. Statistically significant differences between the PBB and comparison groups for each variable are indicated by **p* < .1, ***p* < .05, and ****p* < .01.

Empirical Analysis: Research Question 1

Table 8 reports the average performance for all measures for both groups in the pre- and postimplementation time periods. The performance for each of the population health measures is measured as a percent of patients receiving the appropriate care, as a function of the total number of patients eligible for the indicated care. For example, the mammography performance measure is the total number of patients at facility j who received a mammogram within the preceding 24 months, divided by the total number of patients at facility j who were eligible to receive a mammography.

The clinical safety measure is the CLABSI SIR, which is calculated as the total number of observed CLABSI events per time period divided by predicted number of CLABSI events, based on nationwide data collected by the CDC from over 25,000 hospitals participating in the NHSN. It is important to note that the sample size for the CLABSI SIR measure is substantially smaller than the other measures. This is due to the inclusion criteria defined by the CDC, which states that CLABSI rates can be reported "in any inpatient location where denominator data can be collected, which can include critical/intensive care units (ICU), specialty care areas (SCA), neonatal units including neonatal intensive care units (NICUs), step down units, wards, and long term care units" (CDC, 2020, p.3).

Table 8

Pre/Post Comparison of Quality Performance for All Dependent Variable Measures, 2004-2018

		Group			
		Comparison		PBB	
	Sample size	Pre	Post	Pre	Post
Population health (% performance)[a]					
Mammography	323	74.52	71.90	59.00	66.29
Cervical cancer screening	342	87.05	86.29	83.00	85.85
Diabetes-A1C screening	311	85.45	86.23	79.28	82.20
Diabetes-A1C control	311	72.28	73.15	65.08	68.55
Diabetes-LDL control	311	47.36	48.75	39.80	42.18
Asthma care	334	85.49	96.61	84.20	95.76
Well child visits	267	58.39	77.86	66.07	80.78
Low back pain	358	76.75	81.63	69.33	77.82
Pharyngitis-Appropriate testing	261	77.58	82.92	73.56	80.27
URI-Appropriate treatment	266	93.82	94.85	91.23	94.15
Mental health-7-day follow-up	323	73.37	73.77	76.63	77.22
Mental health-30-day follow-up	323	85.03	85.90	87.28	88.16
Clinical safety[b]					
CLABSI SIR	15	0.81	0.711	1.683	0.748

Note. CLABSI = central line-associated bloodstream infection; SIR = standardized infection ratio.
[a]Population health performance measures are the percent of target patient population receiving the appropriate care.
[b]Clinical safety performance measure is the CLABSI SIR, which is the number of observed infections divided by the number of predicted infections, based on nationwide CDC data.

Mammography

Figure 7 depicts mammography performance in the PBB facilities and comparison facilities over time during the study period. This figure and ones like it for subsequent performance indicators are used to visually inspect trends in the performance measure before the implementation of PBB for the treatment and comparison groups given the assumption of parallel trends in DID analysis (McKenzie, 2020). Generally, the trends for mammography in both the comparison and treatment groups appear similar in the preperiod based on the data in Figure 7.

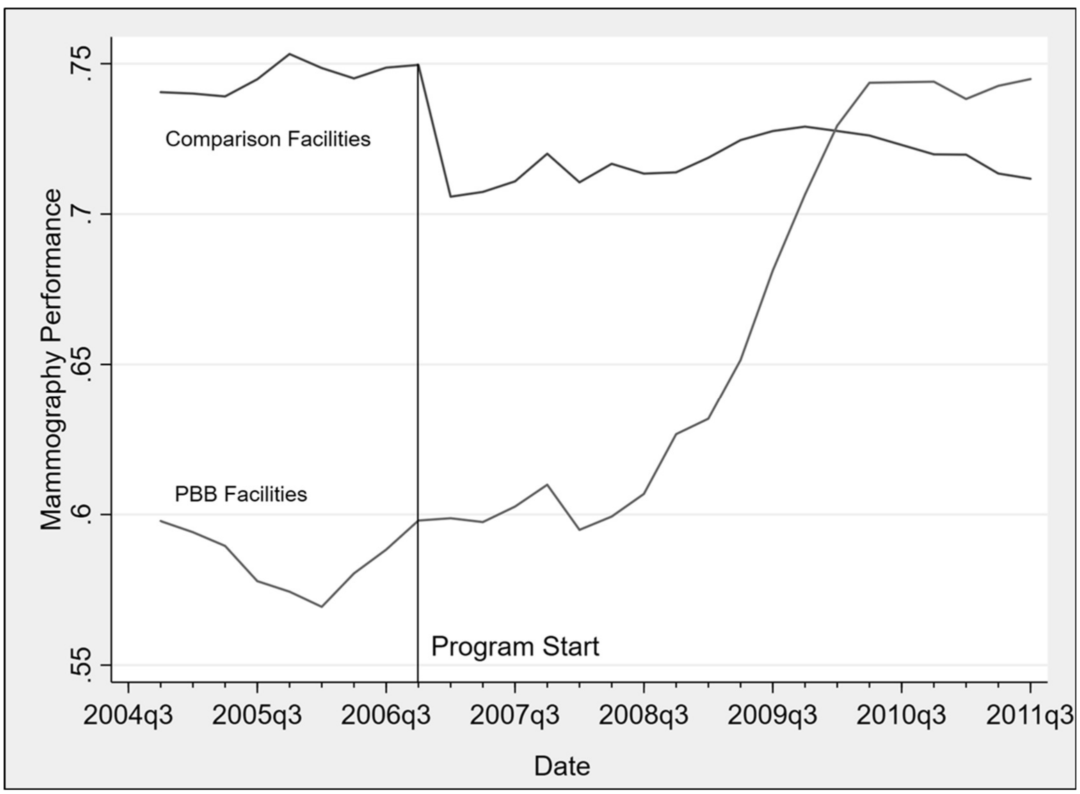

Figure 7. Comparison of mammography performance trends over time (2004-2011). Mammography performance is the percent of eligible patients seen at the facility with mammography within the last 24 months.

Table 9 presents the results generated from a fixed effect Poisson regression with mammography as the dependent variable. Performance is defined as the percentage of eligible patients at each facility who have received mammograms in the previous 24 months (NCQA, 2019).

In support of Hypothesis 1, the coefficient for PBB*Post was significant and positive. These results indicate that the change in performance for the Mammography HEDIS measure for the facilities participating in PBB in the postimplementation period was higher than the comparison facilities (coefficient = 0.069, $p < .001$), controlling for other variables in the model. The acuity and sex variables were not significantly associated with performance on the mammography measure, whereas the age (coefficient = 0.007, $p < .001$), and DHP (coefficient = 0.075, $p < .05$), variables were both significant.

Table 9

Results of Poisson Regression for Mammography Measure

Variable		Coefficient		Robust *SE*
Key independent variables				
	Post*PBB	0.069	***	0.012
	Post	-0.063	***	0.009
Control variables				
	DHP (logged)	0.075	**	0.035
	Acuity (00s)	-0.001		0.002
	Age	0.007	***	0.002
	Sex	0.074		0.079

Note. PBB = performance-based budgeting. DHP = inflation-adjusted, logged value of DHP total budget. SE= Standard Error. Post*PBB = facilities participating in PBB in the postimplementation period. Post = all facilities (comparison and treatment group) in the postimplementation period. Coefficient = parameter estimate for the continuous proportional change in performance for a one-unit increase in the variable, or relative to the base of comparison facilities in the preperiod. Coefficient for the DHP variable is an elasticity.
*$p < .1$. **$p < .05$. ***$p < .01$.

Cervical Cancer Screening

Figure 8 depicts cervical cancer screening performance in the PBB facilities and comparison facilities over time during the study period. Performance on this measure is indicated by the percentage of eligible enrollees at each facility with cervical cancer screening within the previous 36 months (NCQA, 2019). The trends for performance in cervical cancer screening appear to be similar for the treatment and comparison groups in the pre-intervention period, which suggests that parallel trends are likely present.

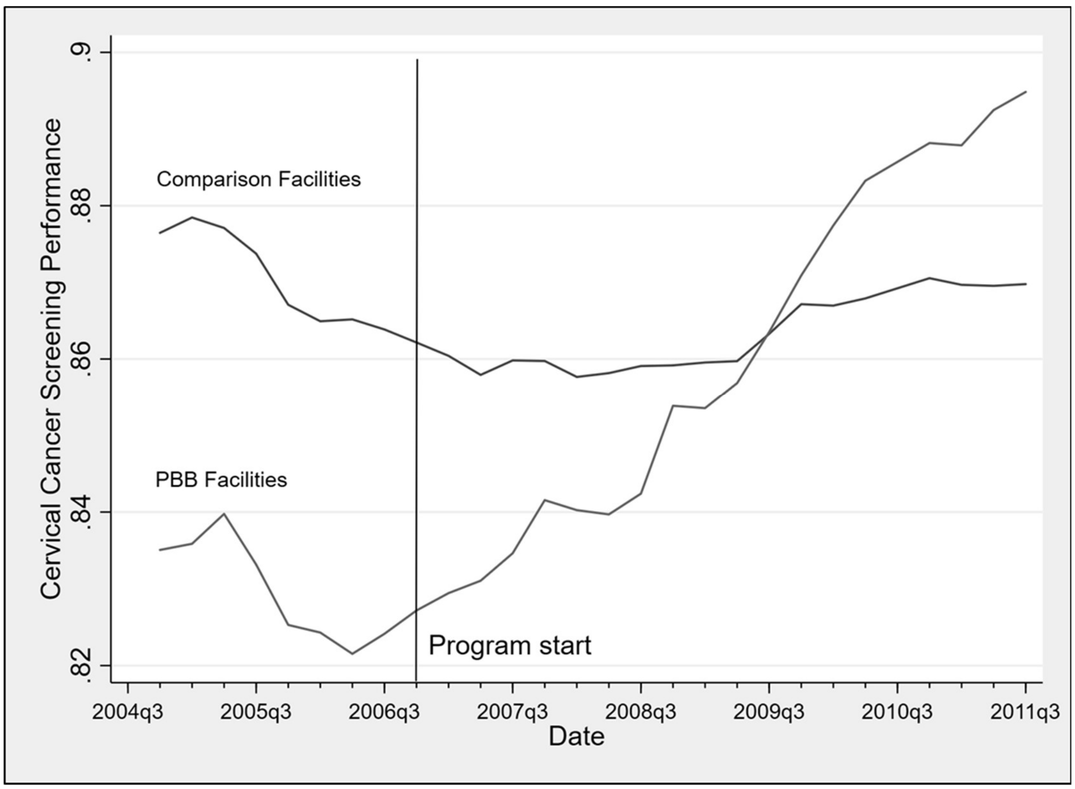

Figure 8. Cervical cancer screening performance over time (2004-2011).

Table 10 presents the results generated from a fixed effect Poisson regression with cervical cancer screening as the dependent variable. In support of Hypothesis 1, the coefficient for the Post*PBB variable was significant and positive. This indicates that the change in performance for the HEDIS cervical cancer screening measure was higher for the facilities participating in PBB in the postimplementation period, relative to comparison facilities, and controlling for other variables in the model (coefficient = 0.0336, $p < .001$). The size of the DHP budget (coefficient: 0.0563; $p < .01$) and acuity (coefficient: -0.002; $p < .01$) were also significantly associated with performance, although the effect size for acuity was small. The age and gender variables for this model were not significant.

Table 10

Results of Poisson Regression for Cervical Cancer Screening Measure

Variable	Coefficient		Robust *SE*
Key independent variables			
Post*PBB	0.0336	***	0.0055
Post	-0.0226	***	0.0031
Control variables			
DHP (logged)	0.0563	***	0.0197
Acuity	-0.002	***	8.52E-04
Age	-0.0012		0.0015
Sex	-0.0404		0.0384

Note. PBB = performance-based budgeting. DHP = inflation-adjusted, logged value of DHP total budget. SE= Standard Error. Post*PBB = facilities participating in PBB in the postimplementation period. Post = all facilities (comparison and treatment group) in the postimplementation period. Coefficient = parameter estimate for the continuous proportional change in performance for a one-unit increase in the variable or relative to the base of comparison facilities in the preperiod. Coefficient for the DHP variable is an elasticity.
*$p < .1$. **$p < .05$. ***$p < .01$.

Diabetes A1C Screening

Figure 9 compares the performance of PBB facilities to the performance of comparison facilities on the diabetes A1C screening measure over time. Performance is indicated by the percentage of eligible patients at each facility with diabetes who have had A1C testing within the previous 12 months (NCQA, 2019). The treatment and comparison groups show similar performance trends in the pre-implementation period with the slope of performance improvement appearing to be higher for the PBB facilities after initiation of the PBAM program in 2006.

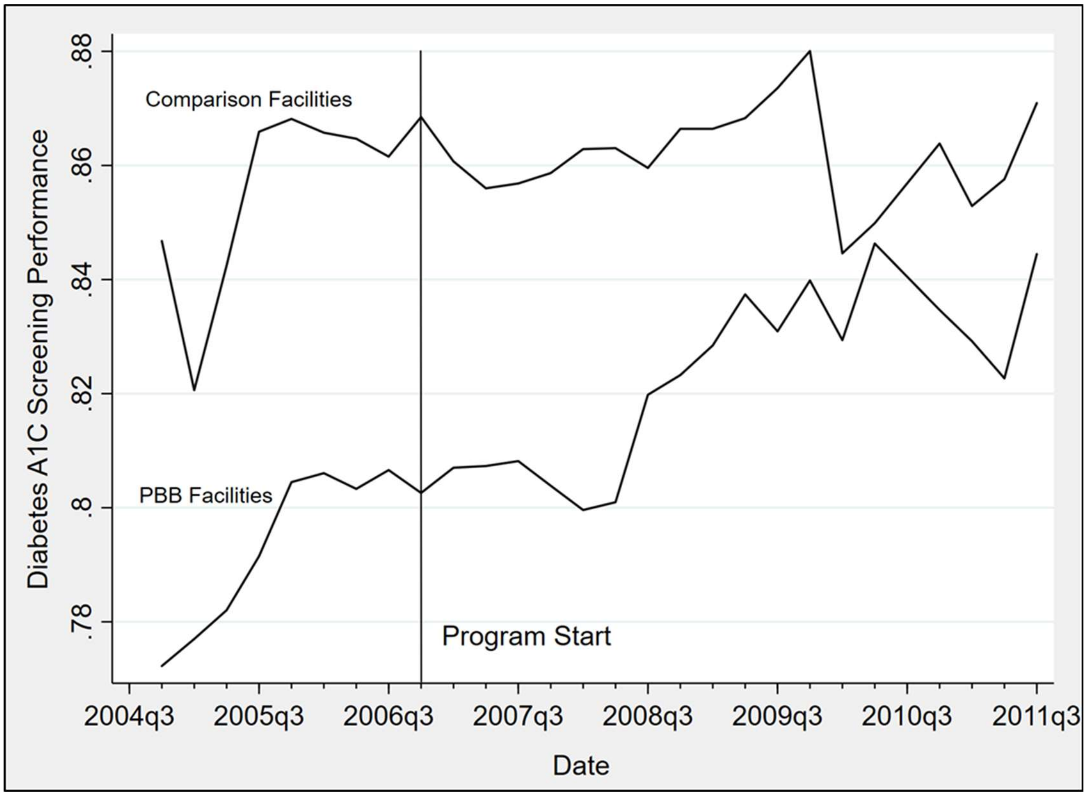

Figure 9. Diabetes A1C screening performance over time (2004-2011).

Table 11 presents the results generated from a fixed effect Poisson regression with diabetes A1C screening as the dependent variable. These results provide some support for Hypothesis 1. Participation in PBB was associated with an increase in performance from the pre-intervention period to the postintervention period in comparison to facilities that did not participate in PBB (coefficient: 0.0175; $p < .001$). An increase in the DHP budget was also associated with an increase in performance (coefficient: 0.1023; $p < .001$), whereas the age, acuity, and gender variables were not significant in this model.

Table 11
Results of Poisson Regression for Diabetes A1C Screening Measure

Variable	Coefficient		Robust *SE*
Key independent variables			
Post*PBB	0.0175	***	0.0053
Post	-0.0178	***	0.0041
Control variables			
DHP (logged)	0.1023	***	0.0220
Acuity (00's)	-0.0011		0.0011
Age	-4E-05		0.0013
Sex	-0.0367		0.0421

Note. PBB = performance-based budgeting. DHP = inflation-adjusted, logged value of DHP total budget. SE= Standard Error. Post*PBB = facilities participating in PBB, in the postimplementation period. Post = all facilities (comparison and treatment group) in the postimplementation period. Coefficient = parameter estimate for the continuous proportional change in performance for a one-unit increase in the variable or relative to the base of comparison facilities in the preperiod. Coefficient for the DHP variable is an elasticity.
$*p < .1$. $**p < .05$. $***p < .01$.

Diabetes A1C Control

Figure 10 presents a comparison of diabetes A1C control performance between the treatment and comparison groups over time. Performance is indicated by the percentage of eligible patients at each facility with diabetes with A1C performance levels at nine or below (NCQA, 2019). Both groups demonstrated a similar pattern of increasing performance through 2005 with performance leveling off through most of 2006. The slope of performance improvement appears to be slightly higher for the PBB facilities in the postimplementation period.

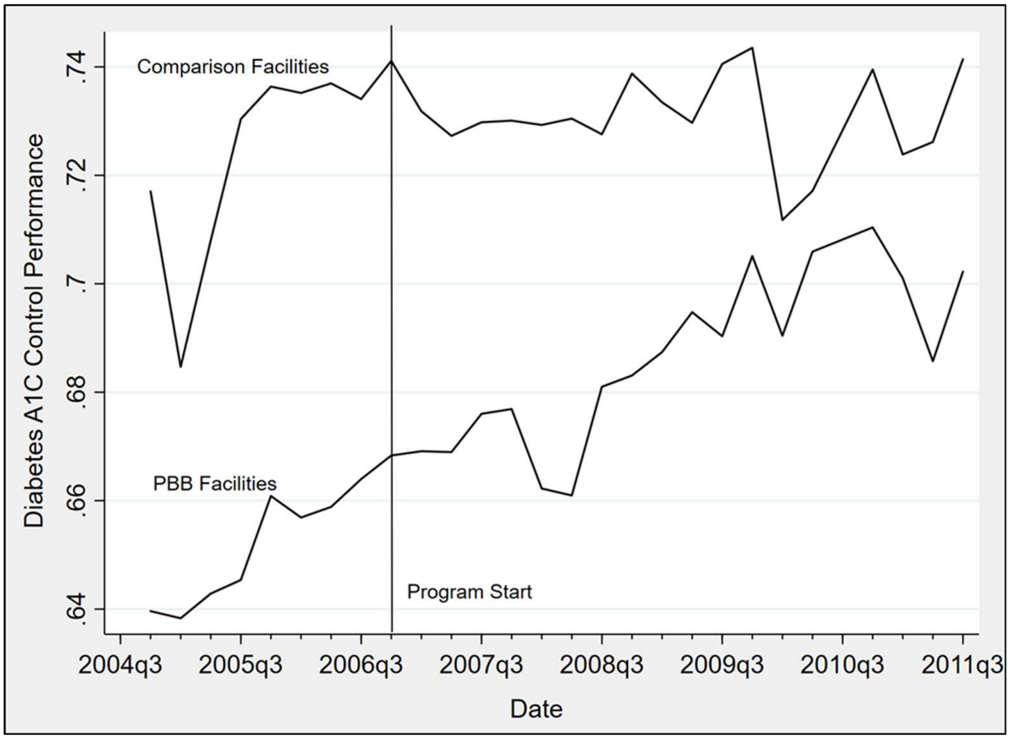

Figure 10. Diabetes A1C control performance over time (2004-2011).

Table 12 presents the results generated from a fixed effect Poisson regression with diabetes A1C control as the dependent variable. The results of this regression provides some support for Hypothesis 1. The coefficient for the Post*PBB variable was significant and positive, indicating that participation in PBB programs in the postimplementation period was associated with higher performance improvement on the diabetes A1C control measure relative to comparison facilities (coefficient: 0.0194; $p < .05$). The size of the DHP budget (coefficient: 0.0933; $p < .001$) and patient population age (coefficient: 0.0038; $p < .05$) were also positively associated with diabetes A1C performance, although the effect size for age was small. The acuity and gender covariates were not significant in this model.

Table 12

Results of Poisson Regression for Diabetes A1C Control Measure

Variable	Coefficient		Robust *SE*
Key independent variables			
Post*PBB	0.0194	**	0.0099
Post	-0.0142	**	0.0068
Control variables			
DHP (logged)	0.0933	***	0.0292
Acuity	-0.0015		0.0017
Age	0.0038	**	0.0018
Sex	-0.0087		0.0609

Note. PBB = performance-based budgeting. DHP = inflation-adjusted, logged value of DHP total budget. SE= Standard Error. Post*PBB = facilities participating in PBB in the postimplementation period. Post = all facilities (comparison and treatment group) in the postimplementation period. Coefficient = parameter estimate for the continuous proportional change in performance for a one-unit increase in the variable or relative to the base of comparison facilities in the preperiod. Coefficient for the DHP variable is an elasticity.
*p < .1. **p < .05. ***p < .01.

Diabetes LDL Control

Figure 11 displays performance trends over time for both groups on the diabetes LDL control measure. Performance is indicated by the percentage of eligible patients diagnosed with diabetes at each facility who have an LDL level below 100 (NCQA, 2019). The PBB facilities and comparison facilities demonstrate a similar pattern of performance improvement in both the pre- and postintervention time periods with the level of performance in PBB facilities remaining below comparison facilities for the entire study period. Based on similarities in performance trends between both groups in the pre-intervention period, Figure 11 suggests that the parallel trends assumption is likely not violated, although there are no obvious performance improvements for the PBB group in the postintervention period in these descriptive data.

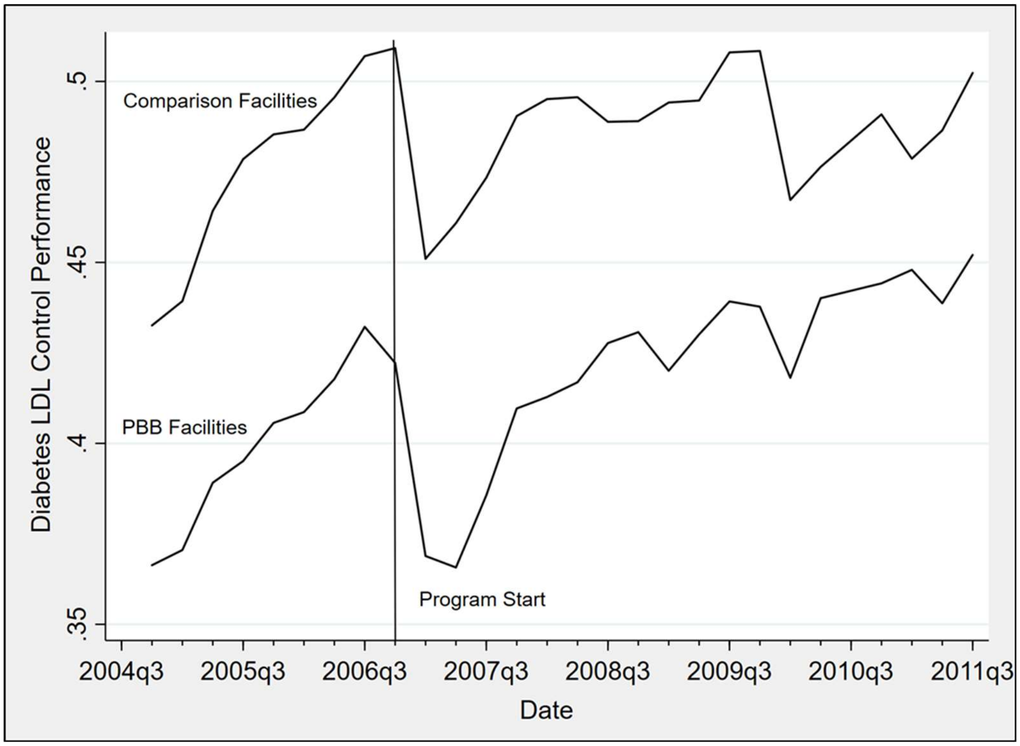

Figure 11. Diabetes LDL control performance over time (2004-2011).

Table 13 presents the results of a fixed effect Poisson regression with diabetes LDL control as the dependent variable. The results fail to support Hypothesis 1. Participation in the PBAM PBB program was not associated with significant performance improvement on the diabetes LDL control measure in the postintervention period relative to comparison facilities. The DHP budget (coefficient: 0.4195; $p < .001$) and age (coefficient: 0.0134; $p < .001$) covariates were both significantly associated with diabetes LDL control performance in this model. However, the effect size for age was small. Patient population acuity and gender were not associated with any effects on performance for this measure.

Table 13

Results of Poisson Regression for Diabetes LDL Control Measure

Variable	Coefficient		Robust *SE*
Key independent variables			
Post*PBB	0.0243		0.0258
Post	-0.0749	***	0.0151
Control variables			
DHP (logged)	0.4195	***	0.0563
Acuity (00s)	-0.0026		0.0054
Age	0.0134	***	0.0050
Sex	0.0028		0.1837

Note. PBB = performance-based budgeting. DHP = inflation-adjusted, logged value of DHP total budget. SE= Standard Error. Post*PBB = facilities participating in PBB in the postimplementation period. Post = all facilities (comparison and treatment group) in the postimplementation period. Coefficient = parameter estimate for the continuous proportional change in performance for a one-unit increase in the variable or relative to the base of comparison facilities in the preperiod. Coefficient for the DHP variable is an elasticity.
**p* < .1. ***p* < .05. ****p* < .01.

Asthma Care

Figure 12 displays performance trends over time for both groups on the asthma care performance measure. Performance on this measure is defined as the percentage of eligible enrollees at each facility with persistent asthma who are prescribed medications considered acceptable as a primary therapy for the long-term control of asthma (NCQA, 2019). The PBB and comparison facilities demonstrated similar patterns in performance throughout the study period. Figure 12 highlights a significant and abrupt change in performance for both groups, occurring between the second and third quarters of 2005. This abrupt change is suggestive of confounding instrumentation effects, possibly related to a change in how performance was measured and recorded.

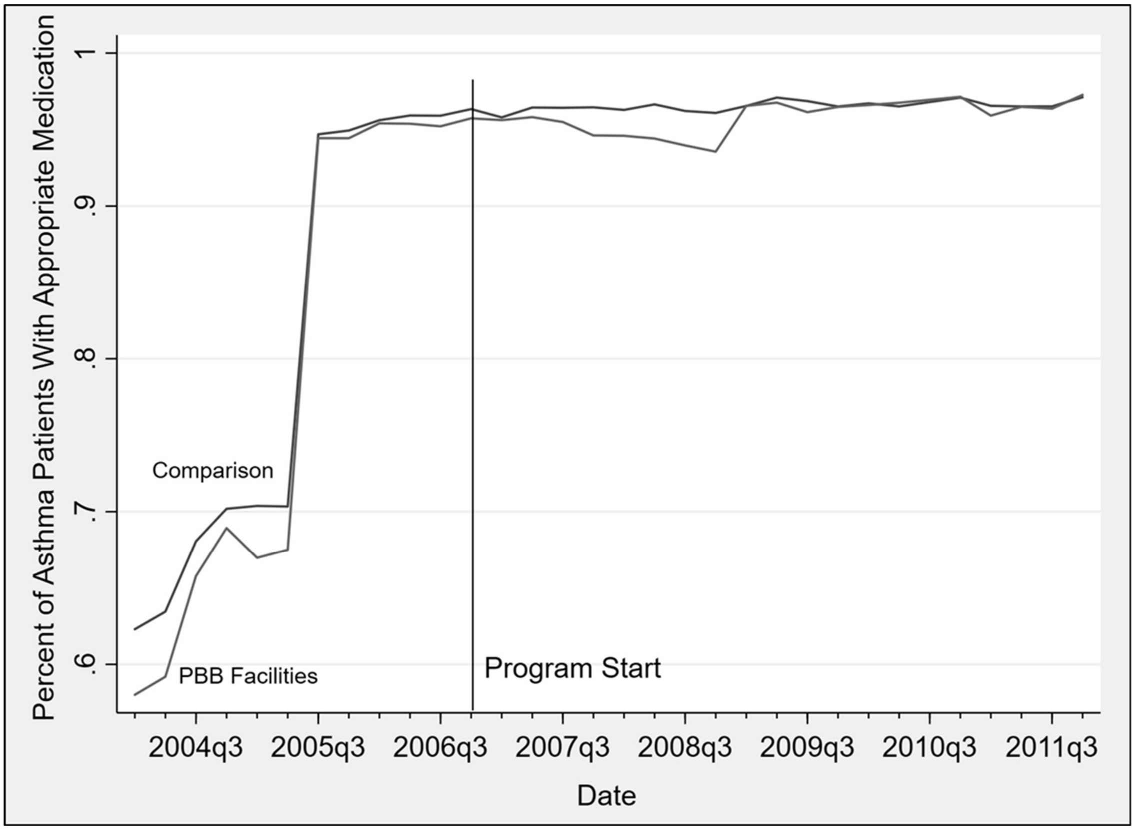

Figure 12. Asthma care performance over time (2004-2011).

Based on Figure 12, the abrupt change in performance occurring in 2005 appears to have impacted both groups similarly. However, since the exact cause of the anomaly could not be confirmed, observations prior to the third quarter of 2005 were dropped from the analysis. Dropping these observations helped to provide a more consistent measurement of pre-implementation performance trends, though the preperiod was shortened. The trend analysis from this abbreviated observation period is presented in Figure 13. It appears that the performance trends prior to the implementation of PBAM incentives were similar for both groups, suggesting that the parallel trends assumption was likely not violated.

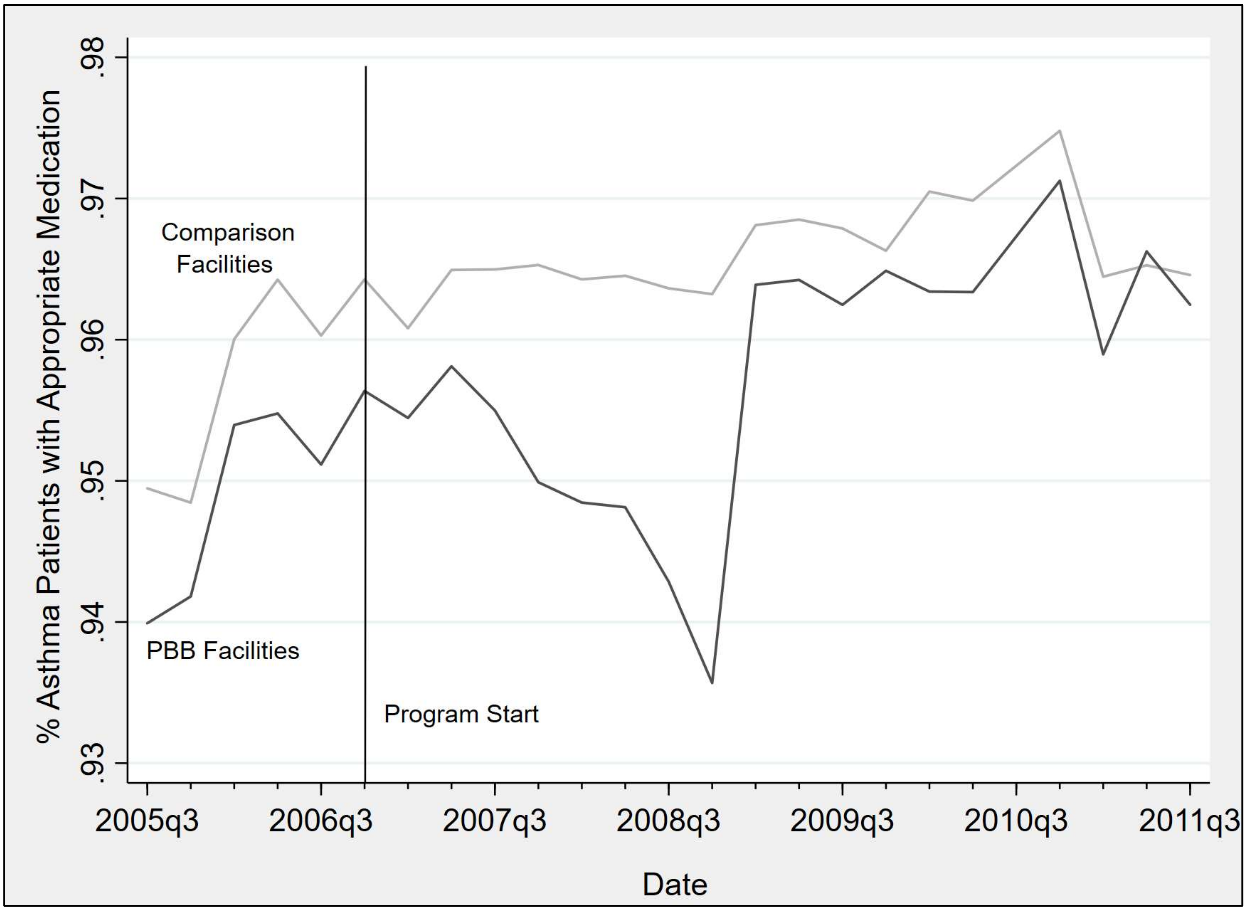

Figure 13. Asthma care performance over time (third quarter 2005-2011).

Table 14 presents the results of a fixed effect Poisson regression with asthma care as the dependent variable. The results do not support Hypothesis 1. The coefficient for the Post*PBB variable was negative and significant, indicating that performance improvement for PBB facilities in the postimplementation period was lower than comparison facilities, controlling for all other model variables (coefficient: -0.0042; $p < .10$). However, it is important to note that this effect size was small. The DHP control variable was significant and positive (coefficient: 0.0465; $p < .001$), while the other model covariates were not significant.

Table 14

Results of Poisson Regression for Asthma Care Performance

Variable	Coefficient		Robust *SE*
Key independent variables			
Post*PBB	-0.0042	*	0.0024
Post	0.0020		0.0021
Control variables			
DHP (logged)	0.0465	***	0.0083
Acuity (00s)	0.0000		0.0004
Age	-0.0003		0.0006
Sex	-0.0016		0.0185

Note. PBB = performance-based budgeting. DHP = inflation-adjusted, logged value of DHP total budget. SE= Standard Error. Post*PBB = facilities participating in PBB in the postimplementation period. Post = all facilities (comparison and treatment group) in the postimplementation period. Coefficient = parameter estimate for the continuous proportional change in performance for a one-unit increase in the variable or relative to the base of comparison facilities in the preperiod. Coefficient for the DHP variable is an elasticity.
*p < .1. **p < .05. ***p < .01.

Well Child Visits

Figure 14 depicts performance on well child visits in the PBB facilities and comparison facilities over time during the study period. Performance is defined by HEDIS criteria for the percentage of eligible children at each facility with six or more well child visits in the first 15 months of life (NCQA, 2019). Both the PBB and comparison groups demonstrate a similar pattern of increasing performance in the pre-implementation period. Based on similar performance improvement trends in the pre-implementation period, the parallel trends assumption seems reasonable for this model.

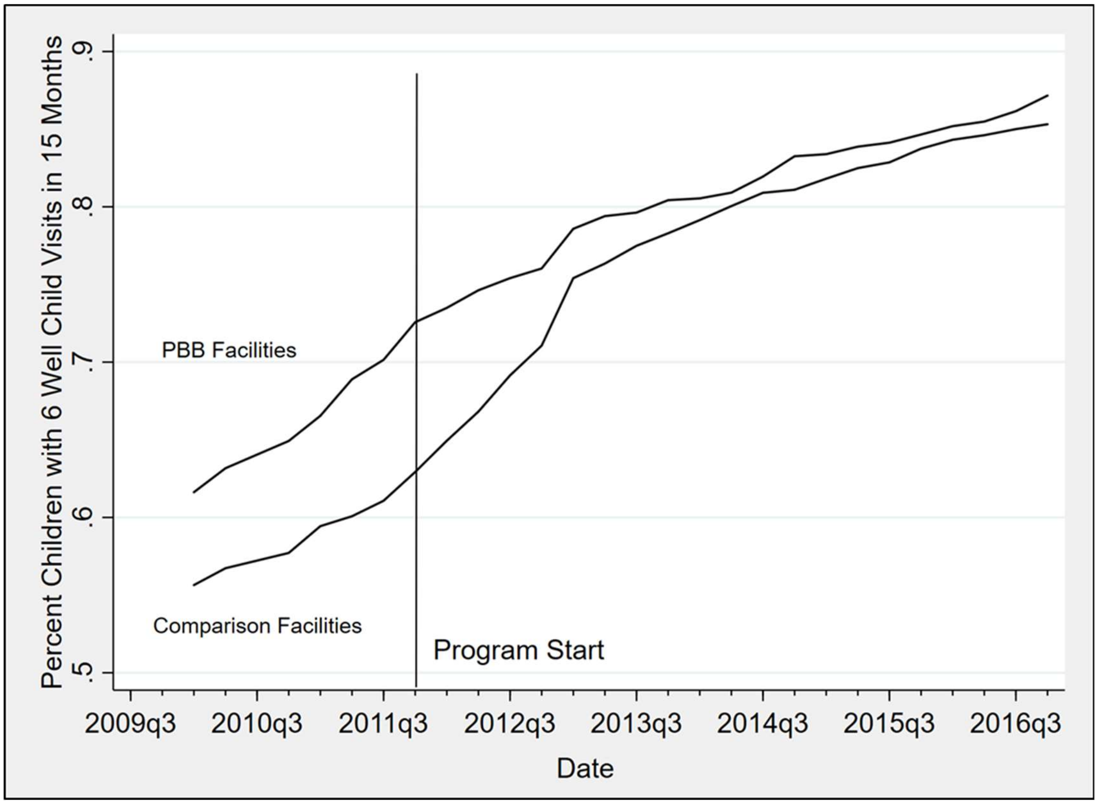

Figure 14. Performance on well child visits over time (2010-2016).

Table 15 presents the results of a fixed effect Poisson regression with well child visits as the dependent variable. The results do not support Hypothesis 1, as the coefficient for the Post*PBB variable is significant and negative. Though both groups exhibited significant performance improvement throughout the study period (coefficient: 0.1713; $p < .001$), the rate of improvement in the pre- and postimplementation periods was slightly lower for the PBB group relative to the comparison group (coefficient: -0.0663; $p < .001$).

Table 15

Results of Poisson Regression for Well Child Performance Measure

Variable	Coefficient		Robust *SE*
Key independent variables			
Post*PBB	-0.0663	***	0.0294
Post	0.1713	***	0.0178
Control variables			
DHP (logged)	-0.9159	***	0.0841
Provider discontinuity	-0.0317		0.0643
Government shutdown	-0.0374	***	0.0044
Acuity	-0.0390	***	0.0048
Age	0.0135	**	0.0059
Sex	0.1028		0.1574

Note. PBB = performance-based budgeting. DHP = inflation-adjusted, logged value of DHP total budget. SE= Standard Error. Post*PBB = facilities participating in PBB in the postimplementation period. Post = all facilities (comparison and treatment group) in the postimplementation period. Coefficient = parameter estimate for the continuous proportional change in performance for a one-unit increase in the variable or relative to the base of comparison facilities in the preperiod. Coefficient for the DHP variable is an elasticity.
**p* < .1. ***p* < .05. ****p* < .01.

Low Back Pain

Figure 15 depicts performance on the low back pain measure in the PBB facilities relative to the comparison facilities over time. Performance is indicated by the percentage of enrollees at each facility with a primary diagnosis of low back pain who did *not* have an imaging study (plan x-ray, MRI, CT scan) within 28 days of diagnosis (NCQA, 2019). The slopes of performance trends are similar for the PBB and comparison facilities in the pre-implementation period, suggesting that the parallel trends are present.

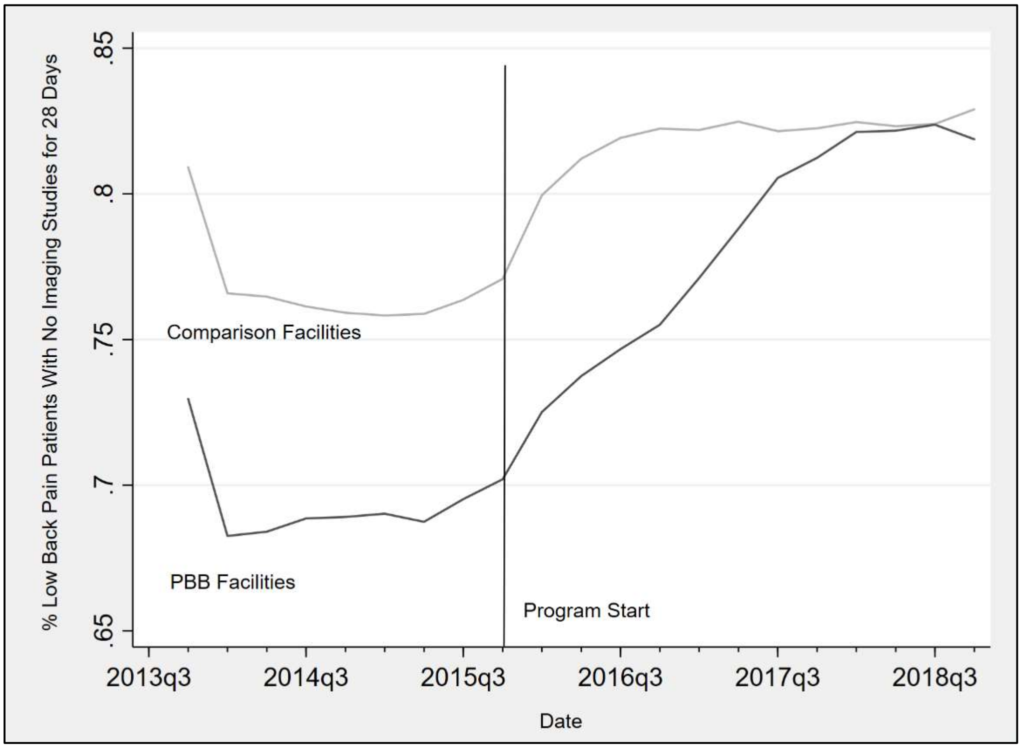

Figure 15. Performance on low back pain measure over time (2013-2018).

Table 16 presents the results of a fixed effect Poisson regression with the HEDIS low back pain measure as the dependent variable. The results of this regression support Hypothesis 1. Both groups exhibited an overall trend of improvement in performance in the postperiod (coefficient: 0.0308; $p < .001$). However, the coefficient for the Post*PBB variable was both significant and positive, indicating that participation in the IRIS PBB program was associated with a higher degree of performance increase in the postperiod relative to comparison facilities (coefficient: 0.0533; $p < .001$). The DHP (coefficient: -0.7524; $p < .001$), government shutdown (coefficient: 0.0281; $p < .001$), and acuity (coefficient: 0.0364; $p < .001$) covariates also yielded statistically significant results, although the magnitude of the effect for acuity was small. The provider discontinuity, age, and sex covariates did not yield statistically significant results for this model.

Table 16

Results of Poisson Regression for Low Back Pain Performance Measure

Variable	Coefficient		Robust *SE*
Key independent variables			
Post*PBB	0.0533	***	0.0086
Post	0.0308	***	0.0058
Control variables			
DHP (logged)	-0.7524	***	0.1071
Provider discontinuity	0.0226		0.0327
Government shutdown	0.0281	***	0.0027
Acuity	0.0364	***	0.0094
Age	0.0020		0.0014
Sex	-0.0439		0.0818

Note. PBB = performance-based budgeting. DHP = inflation-adjusted, logged value of DHP total budget. SE= Standard Error. Post*PBB = facilities participating in PBB in the postimplementation period. Post = all facilities (comparison and treatment group) in the postimplementation period. Coefficient = parameter estimate for the continuous proportional change in performance for a one-unit increase in the variable or relative to the base of comparison facilities in the preperiod. Coefficient for the DHP variable is an elasticity.
*$p < .1$. **$p < .05$. ***$p < .01$.

Appropriate Testing for Pharyngitis

Figure 16 presents a comparison of performance on the appropriate treatment of pharyngitis measure between the treatment and comparison groups over time. Performance is defined by NCQA as the percentage of children 2-18 years of age at each facility who were "diagnosed with pharyngitis, dispensed an antibiotic and received a group A streptococcus test for the episode. A higher rate represents better performance (i.e., appropriate testing)" (NCQA, 2019). The PBB group and comparison group exhibited similar patterns of performance in the pre-implementation period. A visual inspection of performance over time suggests that the parallel trends assumption is reasonable in this case. Though PBB facilities exhibited overall lower levels of performance, it appears that performance rose more sharply for the PBB group after incentives were introduced through the IRIS program in 2017. By the end of 2018, performance in PBB facilities exceeded performance in comparison facilities.

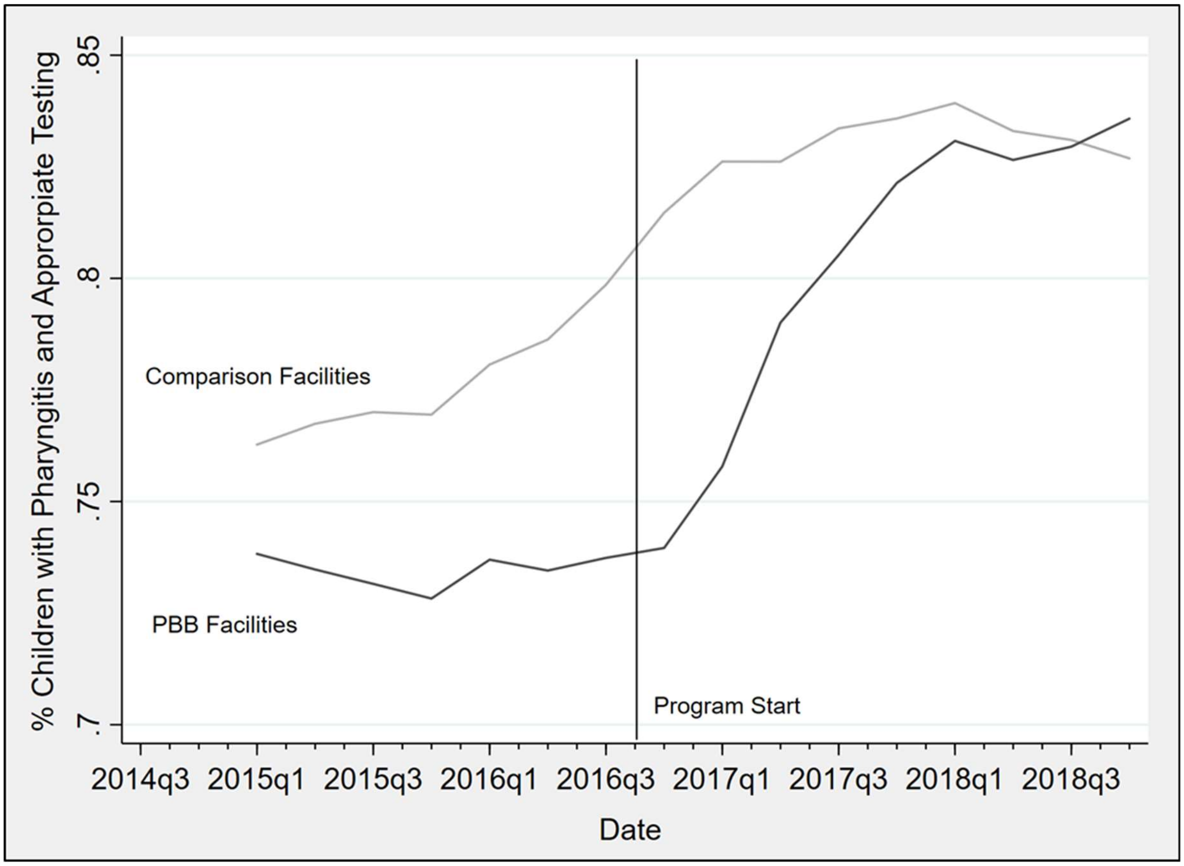

Figure 16. Performance on pharyngitis performance measure over time (2015-2018).

Table 17 presents the results of a fixed effect Poisson regression with the HEDIS pharyngitis measure as the dependent variable. The coefficient for the Post*PBB variable is significant and positive, which supports Hypothesis 1. Both groups exhibited an overall trend of performance improvement in the postperiod (coefficient: 0.0370; $p < .001$), but PBB facilities experienced a sharper increase in performance in the postimplementation period, relative to the comparison facilities (coefficient: 0.0337; $p < .10$).

Table 17

Regression Results for Performance Measure: Pharyngitis

Variable	Coefficient		Robust *SE*
Key independent variables			
Post*PBB	0.0337	*	0.0180
Post	0.0370	***	0.0119
Control variables			
DHP (logged)	-0.6383	***	0.1627
Provider discontinuity	-0.0266		0.0426
Government shutdown	0.0009		0.0072
Acuity	0.0906	***	0.0336
Age	0.0017		0.0024
Sex	0.1655		0.1616

Note. PBB = performance-based budgeting. DHP = inflation-adjusted, logged value of DHP total budget. SE= Standard Error. Post*PBB = facilities participating in PBB in the postimplementation period. Post = all facilities (comparison and treatment group) in the postimplementation period. Coefficient = parameter estimate for the continuous proportional change in performance for a one-unit increase in the variable or relative to the base of comparison facilities in the preperiod. Coefficient for the DHP variable is an elasticity.
*$p < .1$. **$p < .05$. ***$p < .01$.

Appropriate Treatment for URI

Figure 17 presents a comparison of performance over time between both groups on the URI treatment measure. Performance is defined by the NCQA as the percentage of children at each facility, “3 months–18 years of age who were given a diagnosis of URI and were not dispensed an antibiotic prescription. A higher rate indicates appropriate treatment of children with URI (i.e., the proportion for whom antibiotics were not prescribed)” (NCQA, 2019). Figure 17 does not appear to show similarity in performance trends between the two groups in the pre-implementation period. It appears that both groups trend improved performance in the pre-implementation period, but performance appears to have risen more sharply in comparison facilities relative to PBB facilities.

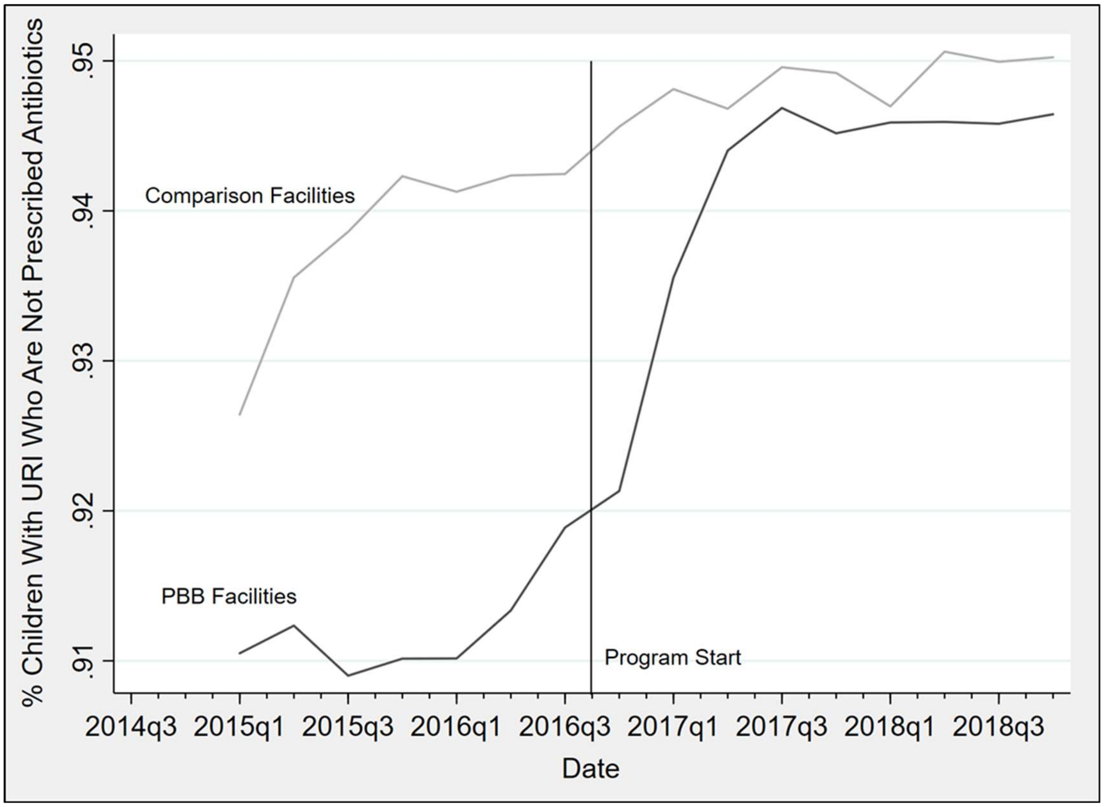

Figure 17. Performance on the URI treatment performance measure over time (2015-2018).

Table 18 presents the results of a fixed effect Poisson regression with the HEDIS URI measure as the dependent variable. The results from this regression support Hypothesis 1. The coefficient for the Post*PBB variable was significant and positive (coefficient: 0.0239; $p < .001$), controlling for other variables in the model. Additionally, the DHP (coefficient: -0.2132; $p < .001$) and acuity (coefficient: 0.0320; $p < .05$) covariates were significant in this model, although the magnitude of the effect for acuity was small. The provider discontinuity, government shutdown, age, and gender covariates were not significant.

Table 18

Results of Poisson Regression for URI Performance Measure

Variable	Coefficient		Robust *SE*
Key independent variables			
Post*PBB	0.0239	***	0.0060
Post	0.0034		0.0024
Control variables			
DHP (logged)	-0.2132	***	0.0452
Provider discontinuity	0.0100		0.0184
Government shutdown	-0.0003		0.0020
Acuity (00's)	0.0320	**	0.0144
Age	0.0010		0.0009
Sex	0.0778		0.0533

Note. PBB = performance-based budgeting. DHP = inflation-adjusted, logged value of DHP total budget. SE= Standard Error. Post*PBB = facilities participating in PBB in the postimplementation period. Post = all facilities (comparison and treatment group) in the postimplementation period. Coefficient = parameter estimate for the continuous proportional change in performance for a one-unit increase in the variable or relative to the base of comparison facilities in the preperiod. Coefficient for the DHP variable is an elasticity.
**p* < .1. ***p* < .05. ****p* < .01.

Seven-Day Follow-Up After Mental Health Hospitalizations

Figure 18 presents a comparison of performance between both groups over time for the mental health hospitalization 7-day follow-up performance measure. Performance is defined as the "percent of adults and children 6 years of age and older who were hospitalized for treatment of selected mental health disorders and received a follow-up within 7 days of discharge" (NCQA, 2019). Figure 18 shows similar trends of performance between both groups in the pre- and postimplementation periods.

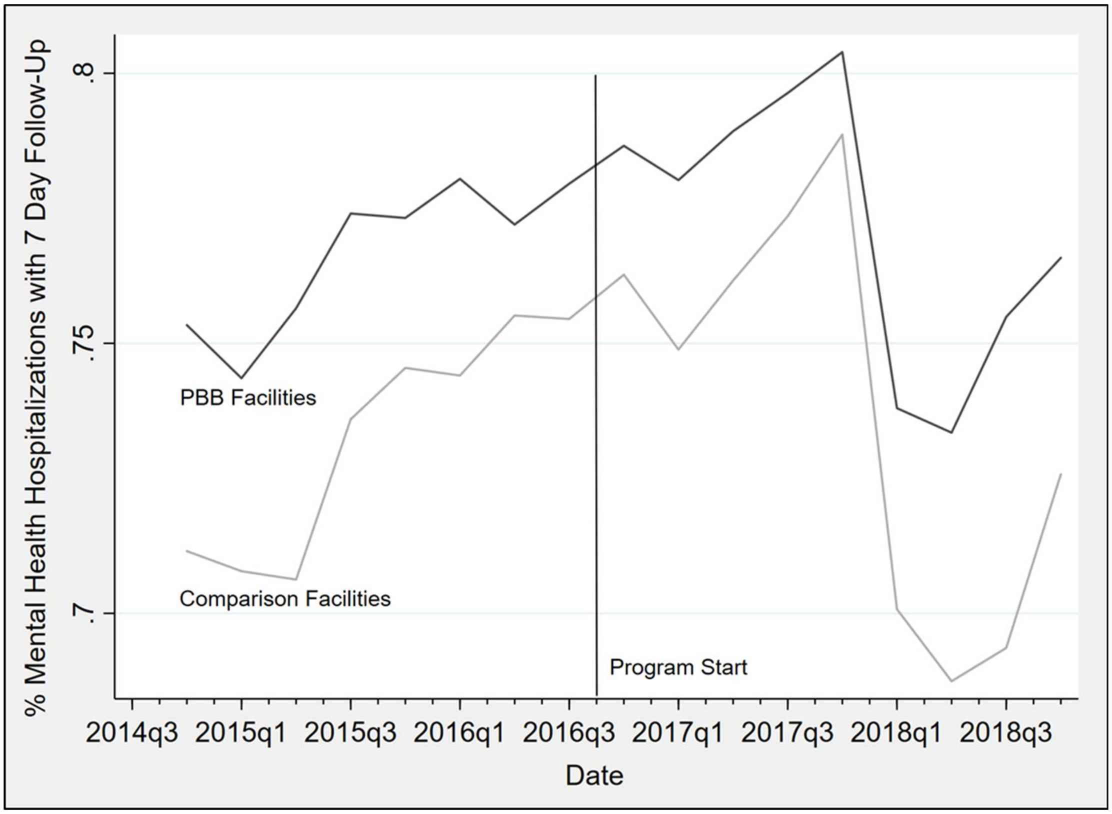

Figure 18. Mental health hospitalization 7-day follow-up measure performance over time (2014-2018).

Table 19 provides the results of a fixed effect Poisson regression with the mental health hospitalization 7-day follow-up measure as the dependent variable. The results of this regression do not support Hypothesis 1. There were no statistically significant differences in performance between the PBB and comparison facilities in the period after the IRIS incentives were introduced for this measure. The size of the DHP budget (coefficient: -1.1526; $p < .001$), provider discontinuity (coefficient: 0.0313; $p < .10$), presence of a government shutdown (coefficient: -0.1195; $p < .001$), average patient age (coefficient: -0.0036; $p < .10$), and female gender (coefficient: -0.2767; $p < .05$) all had significant effects on performance. The acuity covariate was not significant in this model.

Table 19

Results of Poisson Regression for Mental Health Hospitalization 7-Day Follow-Up

Variable	Coefficient		*SE*
Key independent variables			
Post*PBB	-0.0001		0.0122
Post	0.0212	**	0.0089
Control variables			
DHP (logged)	-1.1526	***	0.1394
Provider discontinuity	0.0313	*	0.0411
Government shutdown	-0.1195	***	0.0095
Acuity	-0.0050		0.0244
Age	-0.0036	*	0.0020
Sex	-0.2767	**	0.1390

Note. PBB = performance-based budgeting. DHP = inflation-adjusted, logged value of DHP total budget. SE= Standard Error. Post*PBB = facilities participating in PBB in the postimplementation period. Post = all facilities (comparison and treatment group) in the postimplementation period. Coefficient = parameter estimate for the continuous proportional change in performance for a one-unit increase in the variable or relative to the base of comparison facilities in the preperiod. Coefficient for the DHP variable is an elasticity.
**p* < .1. ***p* < .05. ****p* < .01.

Thirty-Day Follow-Up After Mental Health Hospitalizations

Figure 19 presents a comparison of performance between the PBB and comparison facilities over time for the mental health hospitalization 30-day follow-up performance measure. Performance is defined as the "percent of adults and children 6 years of age and older who were hospitalized for treatment of selected mental health disorders and received a follow-up within 30 days of discharge" (NCQA, 2019). Figure 19 shows similar trends in performance between both groups throughout the entire study period.

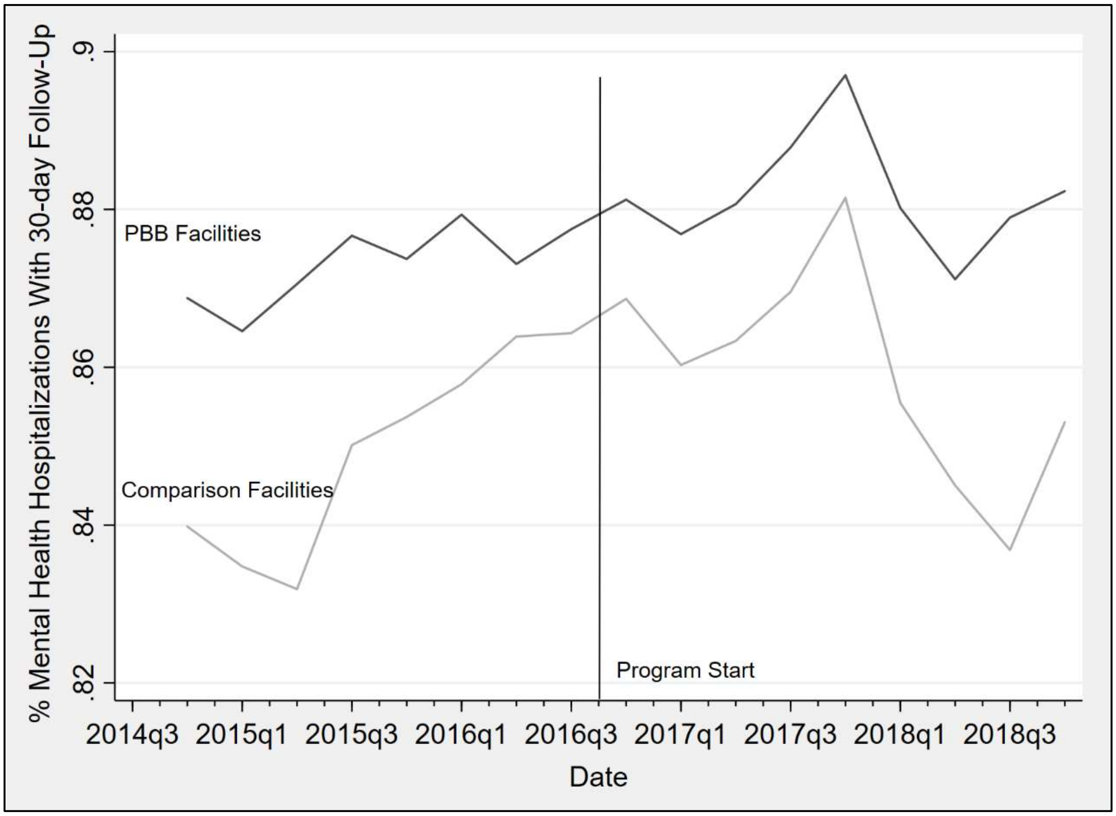

Figure 19. Mental health hospitalization 7-day follow-up measure performance over time (2014-2018).

Table 20 displays the results of a fixed effects Poisson regression with the performance measure for follow-ups within 30 days for mental health hospitalization as the dependent variable. The results do not support Hypothesis 1. The PBB incentives were not associated with a significant change in performance for the PBB facilities relative to the comparison facilities in the postimplementation period. Neither of the groups experienced any significant change in performance between the pre- and postimplementation periods. The size of the DHP budget (coefficient: -0.5109; $p < .001$), presence of a government shutdown (coefficient: -0.0349; $p < .001$), and average patient age (coefficient: -0.0025; $p < .10$) were all associated with significant impacts on performance. The provider discontinuity, acuity, and gender covariates were not significant for this model.

Table 20

Results of Poisson Regression for Mental Health Hospitalization 30-Day Follow-Up

Variable	Coefficient		*SE*
Key independent variables			
Post*PBB	-0.0034		0.0074
Post	0.0089		0.0057
Control variables			
DHP (logged)	-0.5109	***	0.0858
Provider discontinuity	0.0394		0.0241
Government shutdown	-0.0349	***	0.0053
Acuity (00’s)	0.0024		0.0137
Age	-0.0025	*	0.0013
Sex	-0.1338		0.0852

Note. PBB = performance-based budgeting. DHP = inflation-adjusted, logged value of DHP total budget. SE= Standard Error. Post*PBB = facilities participating in PBB in the postimplementation period. Post = all facilities (comparison and treatment group) in the postimplementation period. Coefficient = parameter estimate for the continuous proportional change in performance for a one-unit increase in the variable or relative to the base of comparison facilities in the preperiod. Coefficient for the DHP variable is an elasticity.
*$p < .1$. **$p < .05$. ***$p < .01$.

CLABSI

Figure 20 depicts a trend analysis for clinical safety performance over time for both the PBB and comparison facilities. Performance on clinical safety is defined as the risk-adjusted rate ratio of CLABSIs in hospital ICUs, compared to other participating ICUs in the CDC NHSN Program. The risk-adjusted rate ratio is calculated by dividing the number of observed CLABSIs at each facility by the predicted number of CLABSIs, based on CDC NHSN data for similar facilities (CDC, 2020).

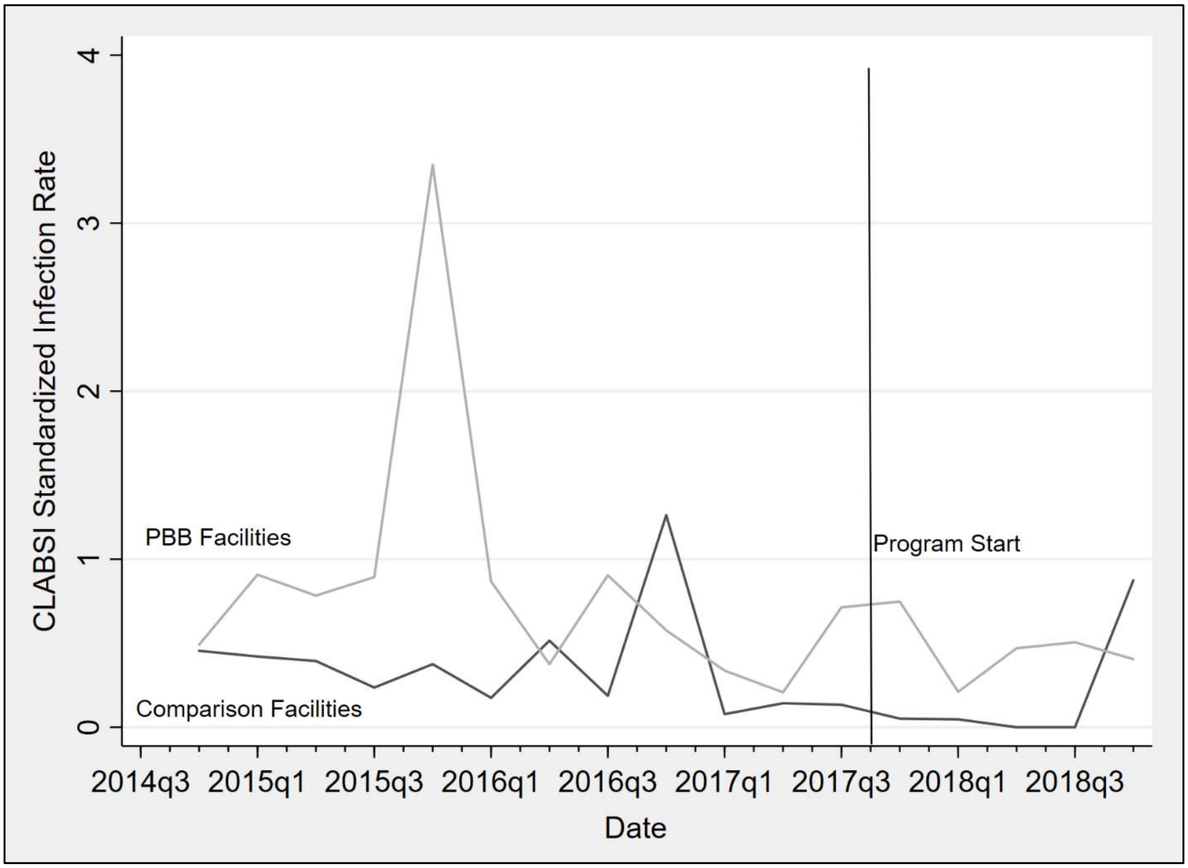

Figure 20. CLABSI performance over time (2014-2018). Standardized infection ratio = observed number of CLABSIs/predicted number of CLABSIs based on CDC NHSN data.

The Y-Axis in Figure 20 displays the average CLABSI SIR, which is the observed count of CLABSIs divided by the CDC-predicted number of CLABSIs. Performance above1 indicates that facilities have a higher rate of CLABSIs than predicted, whereas performance below 1 indicates that facilities have a lower CLABSI rate than predicted. Performance trends on the CLABSI measure were not similar between the PBB and comparison groups prior to the initiation of incentives through the IRIS program. The figure illustrates much variability in rates for both groups over the study period, most likely due to small sample sizes. It is important to note that the sample size for the CLABSI measure (*n* = 15) was significantly lower than the other measures in this study due to the measure-specific inclusion criteria defined by the CDC. The only facilities eligible to report CLABSI rates are inpatient settings that use central lines (CDC, 2020).

Table 21 presents the results of a fixed effect Poisson regression with the CLABSI SIR as the dependent variable. It is important to note that the patient-related variables (age, gender, acuity) were dropped from this analysis because they apply to the outpatient population at each facility. Additionally, this measure is already risk-adjusted using CDC methodology for patient-level risk factors. The provider discontinuity variable was also dropped from this model because it applies primarily to outpatient medical encounters.

Table 21

Results of Poisson Regression for CLABSI SIR Measure

Variable	Coefficient		*SE*
Key independent variables			
Post*PBB	0.9014	**	0.4242
Post	-1.2067	***	0.4394
Control variables			
Government shutdown	-0.0630		0.1831
DHP (logged)	-4.2230		2.6873

Note. PBB = performance-based budgeting. DHP = inflation-adjusted, logged value of DHP total budget. SE= Standard Error. Post*PBB = facilities participating in PBB in the postimplementation period. Post = all facilities (comparison and treatment group) in the postimplementation period. Coefficient = marginal effect of each variable on quality performance.
*$p < .1$. **$p < .05$. ***$p < .01$.

The regression results do not support Hypothesis 1. The coefficient for the post-variable was significant and negative, indicating that CLABSI SIRs dropped for both groups in the postintervention period (coefficient: -1.2067; $p < .001$). Contrary to Hypothesis 1, the coefficient was significant and positive for the Post*PBB variable (coefficient: 0.9014; $p < .05$), indicating that the CLABSI SIR did not drop more for PBB facilities in the postperiod relative to the comparison facilities. However, this result should be interpreted with caution due to small sample sizes and the substantial variability in this measure as depicted in Figure 19.

Summary of Results for Research Question 1

Table 22 presents a summary of the results of all fixed effects Poisson regressions for all performance measures included as dependent variables for Research Question 1. The far right column displays the marginal effect on performance for each measure for which significant effects were observed. In support of Hypothesis 1, PBB was associated with performance gains for the majority of measures: mammography, cervical cancer screening, diabetes A1C screening, diabetes A1C control, low back pain, appropriate testing with pharyngitis diagnosis in children, and appropriate treatment for URI. In contrast to Hypothesis 1, PBB facilities experienced less of an increase in performance in well child visits measure and CLABSI measure in the postimplementation period and a potential decline in performance for the asthma measure relative to comparison facilities. No significant marginal impacts on performance were observed for PBB facilities in the postimplementation period for the diabetes LDL control, follow-up on mental health hospitalization within 7 days, and follow-up on mental health hospitalization within 30 days measures. It is important to note that the results for the CLABSI and URI measures should be interpreted with caution due to the lack of parallel trends observed in a visual inspection of the pre-implementation data.

Table 22

Summary of the Marginal Effects of PBB on Quality Performance

		Marginal effect of PBB on quality performance			
Variable	Sample size (facilities)	Coefficient		*SE*	PBB Marginal Effect[a] (% change in performance)
Population health					
Mammography	323	0.0687	***	0.0118	7.1%
Cervical cancer screening	342	0.0336	***	0.0055	3.4%
Diabetes-A1C screening	311	0.0175	***	0.0053	1.8%
Diabetes-A1C control	311	0.0194	**	0.0099	2.0%
Diabetes-LDL control	311	0.0243		0.0258	--
Asthma care	322	-0.0042	*	0.0024	-0.4%
Well child visits	267	-0.0663	**	0.0294	-6.4%
Low back pain	358	0.0533	***	0.0086	5.5%
Pharyngitis-Appropriate testing	261	0.0337	*	0.0180	3.4%
URI-Appropriate treatment	266	0.0239	***	0.0060	2.4%
Mental health-7-day follow-up	323	-0.0001		0.0122	--
Mental health-30-day follow-up	323	-0.0034		0.0074	--
Clinical safety					
CLABSI	15	0.9014	**	0.4242	

Note. Coefficient is the interaction of PBB in the postimplementation period. It represents the continuous proportional change in performance for each metric in the PBB facilities relative to the comparison facilities.
[a]Only significant results listed. Marginal effect (% simple change in performance) is calculated as [exp(coeff)-1]*100.
**p* < .1. ***p* < .05. ****p* < .01.

It is important to note that some of the control variables were omitted in the analysis for some of the indicators. For the measures that contained observations prior to 2007, the provider discontinuity variable was omitted due to data availability. The provider discontinuity variable requires the use of the National Physician Identifier data, which was not reliably recorded in encounter data until after 2007. Additionally, the analyses for some of the measures did not contain the government shutdown variable. If no government shutdowns occurred during the observation period for a given performance indicator, this covariate was omitted due to collinearity.

Results of Sensitivity Analysis

Several sensitivity analyses were conducted in order to determine the robustness of the results for Research Question 1. Table 23 presents the first group of sensitivity analyses, which tested for anticipatory effects in the 6-month period immediately preceding the start of each set of incentives. Sensitivity analysis 1a (center columns) demonstrates the effect of dropping all observations occurring in the 6-month period prior to the start of incentives. Sensitivity analysis 1b creates a dichotomous variable indicating observations in the 6-month period immediately prior to the start of incentives. The far right columns represent the interaction of PBB in the 6-month anticipatory period. A significant, positive coefficient suggests that anticipatory effects may be present. As an exception, a significant, *negative* coefficient suggests anticipatory effects for measures in which a decrease is desirable (such as the CLABSI measure).

Table 23

Sensitivity Analysis Testing for Anticipatory Effects

Variable	Original results			Sensitivity analysis 1a: 6-month anticipatory period dropped			Sensitivity analysis 1b: Test for significant anticipatory effects		
	Coefficient[a] Post*PBB		*SE*	Coefficient[a] Post*PBB		*SE*	Coefficient[b] Ant*PBB		*SE*
Performance measure									
Mammography	0.0687	***	0.0118	0.6814	***	0.0125	-0.0002		0.0069
Cervical cancer screening	0.0336	***	0.0055	0.0338	***	0.0058	-0.0010		0.0028
Diabetes-A1C screening	0.0175	***	0.0053	0.0178	***	0.0060	0.0011		0.0069
Diabetes-A1C control	0.0194	**	0.0099	0.1791		0.0116	-0.0055		0.0104
Diabetes-LDL control	0.0243		0.0258	0.0204		0.0270	-0.0101		0.0125
Asthma care	-0.0042	*	0.0024	-0.0062	**	0.0028	-0.0049	**	0.0022
Well child visits	-0.0663	**	0.0294	-0.0542	*	0.0314	0.0387	**	0.0170
Low back pain	0.0533	***	0.0086	0.0531	***	0.0093	0.0002	**	0.0076
Pharyngitis-Appropriate testing	0.0337	*	0.0180	0.0286		0.0201	-0.0149		0.0131
URI-Appropriate treatment	0.0239	***	0.0060	0.0224	***	0.0067	-0.0049		0.0041
Mental health-7-day follow-up	-0.0001		0.0122	-0.0016		0.0139	-0.0090		0.0118
Mental health-30-day follow-up	-0.0034		0.0074	-0.0062		0.0083	-0.0124		0.0076
CLABSI SIR	0.9014	**	0.4242	0.8501	**	0.4208	-0.1379		0.2336

Note. PBB = performance-based budgeting; CLABSI = central line-associated bloodstream infection; SIR= standardized infection ratio. SE= Standard Error.

[a]Coefficient is the interaction of PBB in the postimplementation period. It represents the continuous proportional change in performance for each metric in the the PBB facilities relative to the comparison facilities.

[b]Coefficient is the interaction of PBB in the anticipatory period. It represents the continuous proportional change in performance for each metric in the PBB facilities relative to the comparison facilities.

**p* < .1. ***p* < .05. ****p* < .01.

Excluding the observations in the 6-month period immediately preceding the start of the incentives did not have a significant impact on the results of the main analysis. Dropping observations in the 6-month anticipatory period produced only minor changes in statistical significance for four measures: diabetes A1C control, asthma care, well child visits, and appropriate testing for children diagnosed with pharyngitis. This minor change in statistical significance may be attributable to a loss of statistical power associated with dropping observations and not interpreted as a change in results. For the second sensitivity analysis (final columns in Table 23), only three measures demonstrated a statistically significant change in performance in the 6-month anticipatory period: well child visits, asthma, and low back pain. For the asthma indicator, measurement issues that led to the elimination of some pre-PBB data periods may be having some effect on subsequent periods given the graph in Figure 13. The reasons for pre-implementation improvement in well child visits and low back pain require further study to assess why these trends were observed.

Table 24 presents the results of the next group of sensitivity analyses, the first of which evaluates the impact of dropping facility-level covariates from the model and the second of which drops facilities from the analysis that operate in MSMs. Sensitivity analysis 2 considers the possibility that the model may be overspecified while sensitivity analysis 3 accounts for the possibility that facilities operating in MSMs may experience spillover effects from local facilities that are not participating in PBB. The coefficients represent the interaction of PBB in the postimplementation time period. In comparison to the original analysis, sensitivity analysis 2 did not produce any significant changes in results. The magnitude of the effect of PBB in the postperiod was similar for both models across all measures, although two measures had a minor change in statistical significance (diabetes A1C control and well child visits).

Table 24

Results of Sensitivity Analyses Testing the Impact of Covariates and MSM

Variable	Original results			Sensitivity analysis 2: All covariates dropped from analysis			Sensitivity analysis 3: All MSM facilities dropped from analysis		
	Coefficient[a] Post*PBB		*SE*	Coefficient[a] Post*PBB		*SE*	Coefficient[a] Post*PBB		*SE*
Performance measure									
Mammography	0.0687	***	0.0118	0.0641	***	0.0121	0.0811	***	0.0151
Cervical Cancer Screening	0.0336	***	0.0055	0.0357	***	0.0052	0.0398	***	0.0075
Diabetes-A1C Screening	0.0175	***	0.0053	0.0179	***	0.0050	0.0192	**	0.0079
Diabetes-A1C Control	0.0194	**	0.0099	0.0172	*	0.0099	0.0226		0.0152
Diabetes-LDL Control	0.0243		0.0258	0.0141		0.0274	0.0435		0.0351
Asthma Care	-0.0042	*	0.0024	-0.0040	*	0.0024	-0.0026		0.0033
Well Child Visits	-0.0663	**	0.0294	-0.0978	***	0.0273	-0.0369		0.0327
Low Back Pain	0.0533	***	0.0086	0.0551	***	0.0086	0.0744	***	0.0106
Pharyngitis-Appropriate testing	0.0337	*	0.0180	0.0335	*	0.0181	0.0352	*	0.0203
URI-Appropriate treatment	0.0239	***	0.0060	0.0242	***	0.0062	0.0356	***	0.0082
Mental Health-7-day follow-up	-0.0001		0.0122	0.0006		0.0126	-0.0137		0.0151
Mental Health-30-day follow-up	-0.0034		0.0074	-0.0042		0.0074	-0.0134		0.0087
CLABSI SIR	0.9014	**	0.4242	0.9471	**	0.4222	0.9686	[b]	0.7554

Note. PBB = performance-based budgeting; CLABSI = central line-associated bloodstream infection; SIR= standardized infection ratio; MSM = multi-service market. SE= Standard Error.
[a]Coefficient is the interaction of PBB in the postimplementation period. It represents the continuous proportional change in performance for each metric in the PBB facilities relative to the comparison facilities.
[b]Sample size dropped to 7.
**p* < .1. ***p* < .05. ****p* < .01.

Sensitivity analysis 3 produced some minor changes in the magnitude of the effect sizes for PBB in the post period. In most cases, dropping MSM facilities from the analysis strengthened the effect sizes for PBB, providing some evidence that facilities in these markets may be experiencing some spillover effects. For some measures, such as the diabetes A1C screening, diabetes A1C control, asthma, and the well child visits measures, dropping MSM facilities produced a minor loss of statistical significance. This may be attributable to a loss of statistical power resulting from dropping over 100 facilities from the analysis. Despite these changes, the overall conclusions drawn from the original analysis remain the same. For most performance measures, there was a significant, positive change in performance among PBB facilities relative to comparison facilities after the start of incentives. The purpose of the fourth sensitivity analysis was to determine if the results were potentially biased by facilities that had very few patients eligible for each type of care or treatment (i.e., a low denominator for each performance measure). The original analysis for Research Question 1 eliminated any observations with fewer than 5 patients in the facility who were eligible for the care or treatment. The same model was estimated for sensitivity analyses 4a and 4b, except facilities were dropped if they had fewer than 10 or 20 patients eligible for the care, respectively. On average, 20 facilities were dropped from the analysis for each measure when the minimum number of eligible patients was changed from 5 to 10. Another 24 facilities were dropped, on average, when the minimum number of eligible patients was changed from 10 to 20.

Table 25 presents the results of the fourth sensitivity analysis. There were no appreciable changes to the coefficients, standard errors, or level of statistical significance for any of the measures as a result of changing the threshold for the minimum number of patients necessary for inclusion in the analysis from 5 to 20.

Table 25

Results of Sensitivity Analyses Testing the Impact of Altering Minimum Patient Thresholds

Variable	Original results: Observations with < 5 eligible patients dropped			Sensitivity analysis 4a: Observations with < 10 eligible patients dropped			Sensitivity analysis 4b: Observations with < 20 eligible patients dropped		
	Coefficient[b] Post*PBB		*SE*	Coefficient[b] Post*PBB		*SE*	Coefficient [b] Post*PBB		*SE*
Performance measure[a]									
Mammography	0.0687	***	0.0118	0.0688	***	0.0118	0.0689	***	0.0118
Cervical cancer screening	0.0336	***	0.0055	0.0336	***	0.0055	0.0335	***	0.0055
Diabetes-A1C screening	0.0175	***	0.0053	0.0175	***	0.0053	0.0179	***	0.0053
Diabetes-A1C control	0.0194	**	0.0099	0.0194	**	0.0099	0.0197	**	0.0100
Diabetes-LDL control	0.0243		0.0258	0.0241		0.0259	0.0241		0.0260
Asthma care	-0.0042	*	0.0024	-0.0043	*	0.0024	-0.0044	*	0.0024
Well child visits	-0.0663	**	0.0294	-0.0659	**	0.0294	-0.0655	**	0.0296
Low back pain	0.0533	***	0.0086	0.0534	***	0.0086	0.0530	***	0.0086
Pharyngitis-Appropriate testing	0.0337	*	0.0180	0.0338	*	0.0180	0.0351	*	0.0183
URI-Appropriate treatment	0.0239	***	0.0060	0.0238	***	0.0060	0.0238	***	0.0060
Mental health-7-day follow-up	-0.0001		0.0122	-0.0043		0.0126	-0.0070		0.0133
Mental health-30-day follow-up	-0.0034		0.0074	-0.0057		0.0076	-0.0067		0.0079

Note. Eligible patients are defined as the number of patients eligible for each type of care or treatment for each performance measure. Analysis were conducted with minimum thresholds of 5, 10, and 20. PBB = Performance-based budgeting. SE= Standard Error.
[a]CLABSI measure omitted from this sensitivity analysis because "eligible patients" is not applicable to this measure.
[b]Coefficient is the interaction of PBB in the postimplementation period. It represents the continuous proportional change in performance for each metric in the PBB facilities relative to the comparison facilities.
*p < .1. **p < .05. ***p < .01.

The fifth sensitivity analysis altered the age variable in the original model. The average patient age variable was replaced by a comprehensive set of dummy variables representing 8 age bands. This enabled distinction between facilities on the age distribution of the patient population, which may have otherwise been obscured by using the average age variable. This new model did not produce any appreciable changes to the study results.

The final sensitivity analysis evaluates the impact of including a comprehensive set of quarterly time dummy variables into the model. The purpose of this analysis is to account for potential seasonal effects or trends in the outcomes. Since the DHP budget changes on a yearly basis, it was dropped from this model due to collinearity with time.

Table 26 presents the results of the sixth sensitivity analysis. Changing the time variable to a comprehensive set of quarterly dummy variables did not have an appreciable effect on the results. For most measures, the effect size was slightly larger but the change was small. For example, the marginal effect in mammography performance for PBB facilities relative to comparison facilities in the postperiod changed from 7.1% improvement to 7.4% improvement. The other measures demonstrated similarly small changes as a result of this change in specification of the time variable.

Table 26

Results of Sensitivity Analysis Changing the Specification of Time Variable

Variable	Original results: Coeff.[a] Post*PBB		Original results: *SE*	Sensitivity analysis 6: Accounting for seasonal trends: Coeff.[a,b] Post*PBB		Sensitivity analysis 6: Accounting for seasonal trends: *SE*
Performance measure						
Mammography	0.0687	***	0.0118	0.0717	***	0.0116
Cervical cancer screening	0.0336	***	0.0055	0.0346	***	0.0055
Diabetes-A1C screening	0.0175	***	0.0053	0.0186	***	0.0054
Diabetes-A1C control	0.0194	**	0.0099	0.0203	**	0.0101
Diabetes-LDL control	0.0243		0.0258	0.0225		0.0260
Asthma care	-0.0042	*	0.0024	-0.0034		0.0023
Well child visits	-0.0663	**	0.0294	-0.0799	***	0.0294
Low back pain	0.0533	***	0.0086	0.0546	***	0.0085
Pharyngitis-Appropriate testing	0.0337	*	0.0180	0.0342	*	0.0181
URI-Appropriate treatment	0.0239	***	0.0060	0.0239	***	0.0060
Mental health-7-day follow-up	-0.0001		0.0122	0.0005		0.0120
Mental health-30-day follow-up	-0.0034		0.0074	-0.0030		0.0073
CLABSI SIR	0.9014	**	0.4242	0.8404	**	0.4190

Note. Sensitivity analysis includes a comprehensive set of quarterly time dummy variables to account for seasonal trends in performance on each measure. PBB = performance-based budgeting; CLABSI = central line-associated bloodstream infection; SIR = standardized infection ratio.
[a]Coefficient is the interaction of PBB in the postimplementation period. It represents the continuous proportional change in performance for each metric in the PBB facilities relative to the comparison facilities.
[b]DHP variable omitted from model due to collinearity.
**p* < .1. ***p* < .05. ****p* < .01.

Empirical Analysis: Research Question 2

Table 27 presents a summary of the results of all fixed effects Poisson regressions for those performance measures with sufficient postimplementation data to be examined as a dependent variable for Research Question 2. As discussed previously, Research Question 2 seeks to determine if trends in performance improvement attributable to PBB are sustained over time in the postimplementation period. This is assessed by breaking the 5-year postimplementation period into two parts, evaluating how the performance trends in the first 10 quarters compare to performance in the second 10 quarters after the start of incentives. The two left columns in Table 27 report the interaction of PBB in the first and second postimplementation periods. The coefficients in each of these columns represent the continuous proportional change in performance for PBB facilities in each period relative to comparison facilities. The final column presents the results of postestimation tests evaluating the difference between the coefficients in the first two columns. Values less than 0.10 indicate a significant change in performance between the first 10 quarters and the second 10 quarters of the postimplementation period, for PBB facilities.

Table 27

Summary of the Regression Results for Research Question 2

	Postperiod 1 Quarters 1-10 after start of incentives		Postperiod 2 Quarters 11-20 after start of incentives		Significance of performance change postperiod 1 to postperiod 1[b]
Variable	Coefficient[a]		Coefficient[a]		Prob > Chi2
Population health					
Mammography	0.0442	***	0.0966	***	0.0000
Cervical cancer screening	0.0214	***	0.0497	***	0.0000
Diabetes-A1C screening	0.0047		0.3206	***	0.0000
Diabetes-A1C control	0.0047		0.0360	***	0.0001
Diabetes-LDL control	-0.0046		0.0529	*	0.0016
Asthma care	-0.0061		-0.0015		0.0383

Note. Measures with less than 20 quarters of post data are excluded.
[a]Coefficient is the interaction of PBB in the postimplementation period(s). It represents the continuous proportional change in performance for each metric in the PBB facilities relative to the comparison facilities.
[b]PBB facilities. Based on chi2 postestimation tests: PBB*Period1-PBB*Period2 = 0.
*$p < .1$. **$p < .05$. ***$p < .01$.

As depicted in Table 27, PBB facilities experienced increased performance relative to comparison facilities from period 1 to period 2 for all measures. For the mammography and cervical screening measures, the change in performance for PBB facilities in both periods is significantly higher than comparison facilities. These findings support Hypothesis 2 by suggesting that performance improved for PBB facilities in the postimplementation periods (relative to comparison facilities) and that performance gains continued throughout the 5-year postimplementation period and in fact grew in the second postimplementation period when compared to the first. For the diabetes A1C screening measure, the diabetes A1C control measure, and the diabetes LDL control measure, performance gains for PBB facilities relative to comparison facilities were not significant until the second 10 quarters of the postimplementation period. This partially supports Hypothesis 2 in that no significant gains were present initially but instead there appears to have been a delayed response that led to performance improvements in quarters 11-20 after the incentives were introduced. The results for the asthma care measure fail to support Hypothesis 2. The change in performance for PBB facilities relative to comparison facilities did not reach statistical significance in either half of the postimplementation period. Even though performance improved for PBB facilities from period 1 to period 2, there was never a statistically significant, positive change in performance for PBB facilities relative to comparison facilities, suggesting that PBB did not have a significant marginal effect for that measure.

Summary

This chapter provided the results of the descriptive analyses, main empirical analyses, and sensitivity analyses that address the study's two research questions. In reference to Research Question 1, the study findings indicate that PBB was associated with increased quality performance on seven out of 13 measures under investigation. PBB facilities had less quality performance improvement relative to comparison facilities on three out of 13 of the measures under investigation. Another three out of 13 measures demonstrated no significant change in performance for PBB facilities relative to comparison facilities. The study findings also suggest that performance improvement attributable to PBB was either sustained or improved for at least 5 years for five out of six measures under investigation.

The next chapter summarizes and interprets the empirical results of this study. It also provides an evaluation of the study's limitations and discusses the practical and policy implications of the research findings. The chapter concludes with a suggestion for relevant future research on PBP in healthcare settings.

Chapter 6: Discussion

PBB is a promising but underresearched approach for encouraging quality improvement in government-funded healthcare facilities. The purpose of this research was to investigate the effects of PBB on performance improvement in U.S. military healthcare facilities. This was accomplished by addressing two research questions, both of which pertained to performance improvement on healthcare quality metrics. The first research question sought to determine if PBB was associated with performance improvement in military healthcare facilities that implemented it versus military healthcare facilities that did not. The second research question sought to determine if quality improvements tied to PBB were sustained over time in those same facilities.

The extant research on PBB and its related approach, pay for performance, were discussed in Chapter 2. Chapter 2 also presented reasons why PBB should be explored as a *special case* of pay for performance and why PBB may produce similar effects on quality. Chapter 2 concluded with an overview of the MHS and the conditions that gave rise to an experimental program using PBB in Army healthcare facilities. Chapter 3 presented a conceptual model that applied the concepts of RDT to explain why PBB might give rise to performance improvement in military healthcare facilities. This conceptual model was applied to generate the study's two hypotheses that PBB would be associated with quality performance improvement in healthcare facilities, and that performance improvement would be sustained for at least 5 years in the presence of performance incentives.

Chapter 4 outlined an empirical method for testing the study's hypotheses and conceptual model. A post-hoc DID approach was proposed to test the impact of PBB in Army healthcare facilities by measuring performance on various quality metrics before and after implementation of incentives, and comparing these performance trends to performance of facilities that did not participate in PBB. Detailed results of that analysis were presented in Chapter 5. Chapter 6 starts with a summary and interpretation of those findings, followed by a discussion of the study's limitations. This chapter concludes with a presentation of the study's implications and future research directions.

Summary of Research Findings

This study applied RDT to examine the relationship between PBB and quality performance improvement in military healthcare facilities. Because military healthcare facilities are dependent upon financial resources from the DHP, if access to those resources are made contingent upon quality performance, RDT predicts that healthcare facilities will improve performance on those metrics that are incentivized. Hypothesis 1 stated that there would be greater performance improvement on incentivized quality metrics in Army healthcare organizations that participated in PBB programs relative to military healthcare organizations in other branches that did not participate in PBB programs, ceteris paribus. The study tested Hypothesis 1 using a DID, Poisson regression analysis to compare performance on a variety of quality measures between PBB and non-PBB facilities, both before and after implementation of PBB incentives. Table 28 summarizes the findings for Research Question 1 and specifies whether the hypothesis was supported for each measure examined.

Table 28

Summary of Research Question 1 Findings Categorized by Performance Metric

Hypothesis	Supported
Research Question 1: What is the impact of performance-based budgeting on quality improvement in U.S. Army healthcare facilities?	
Hypothesis 1: There will be greater performance improvement on incentivized quality metrics in Army healthcare organizations that participated in PBB programs relative to military healthcare organizations in other branches that did not participate in PBB programs, ceteris paribus.	
Measures	
Population health	
Mammography	Yes
Cervical cancer screening	Yes
Diabetes-A1C screening	Yes
Diabetes-A1C control	Yes
Diabetes-LDL control	No
Asthma care	No
Well child visits	No
Low back pain	Yes
Pharyngitis-Appropriate testing	Yes
URI-Appropriate treatment	Yes
Mental health-7-day follow-up	No
Mental health-30-day follow-up	No
Clinical safety	
CLABSI	No

Note. PBB = performance-based budgeting; LDL = low density lipoproteins; URI = upper respiratory infection; CLABSI = central line-associated bloodstream infection.

These findings lend support to Hypothesis 1. PBB was associated with performance improvement for over half of the measures under investigation. Though these findings are specific to PBB in military healthcare facilities, they are generally consistent with the literature on pay for performance. The measures with stronger associations between PBB and performance improvement tended to be process measures such as mammography and cervical cancer screening with weaker effect sizes (or no effect) observed for outcome measures such as diabetes management, CLABSI, and asthma care. Three measures (diabetes LDL control, and 7- and 30-day mental health hospitalization follow-ups) were not associated with any performance effects in PBB facilities after initiation of incentives. It is unclear why this pattern of results emerged for the mental health follow-up measures. It is possible that these measures are more challenging for performance improvement because they require robust systems of case management and communication between teams of healthcare professionals. Clarifying underlying reasons for these findings is beyond the scope of this research, but it is worthy of future investigation.

Contrary to Hypothesis 1, the results for the well child visits performance measure appeared to show a *decrease* in performance for PBB facilities in the period after program incentives were offered. It is unclear why this result was observed, but it may relate to the level of mobility among military families during this period. Since performance on this metric spans 15 months, there is some potential for spillover effects from non-PBB facilities because patients move between military installations. For example, a family might move from an Army facility to a MSM and transfer care to a Navy, Air Force, or civilian facility (or vice versa). Thus, performance from one facility may be *carried over* into performance in a new facility for children between the ages of zero and 15 months.

Additionally, during the study time period, there were a significant number of deployments for Army personnel. When servicemembers deploy, families often move away from military installations for the duration of the deployment. For the families remaining on military installations, there is often a strain on the nondeploying spouse. This can potentially lead families to forgo routine wellness care, particularly if adequate childcare resources are not available for other children in the family in order to support participation in wellness care. This unexpected finding warrants further investigation.

Research Question 2 of this study investigated the effects of PBB over a 5-year postperiod to determine whether performance changes were sustained, with Hypothesis 2 predicting that quality improvements attributable to PBB programs will persist throughout the postperiod after each metric is incentivized. The results of this analysis lend support for Hypothesis 2. Five of six measurements demonstrated statistically significant improvement, relative to comparison facilities, by the second half of the postperiod.

Of the measures for which PBB facilities demonstrated significant improvement in the postperiod, three measures did not achieve that level of improvement until the second half of the postperiod. This suggests that incentives may have a delayed effect on performance in some cases. Since all three of these measures pertained to diabetes management, it is possible that the delayed improvement was related to the relative complexity of performance improvement specific to these measures. In contrast to simpler process measures, diabetes management indicators may take longer to produce an observable improvement, possibly due to the need for more complex and sustained patient engagement. PBB was associated with more immediate effects for the mammography and cervical cancer screening measures, which showed significant improvement in both phases of the postperiod. For both measures, performance in PBB facilities improved in the second half of the postperiod relative to the first. It is possible that the immediate effects observed with these two measures are related to the fact that they are process measures that can be impacted within a single visit per patient.

Table 29 presents a summary of findings for Research Question 2. The only indicator that did not show significant improvement relative to comparison facilities in either period was the asthma measure. It is unclear why this result was observed, but it possible that it may be related to the abbreviated preperiod, which reduced the number of observations for that timeframe. Due to this concern, this result should be interpreted with caution.

Overall, the empirical results suggest that PBB is associated with performance improvement effects for a select set of quality performance measures in military healthcare facilities. Consistent with pay for performance research, the largest effects are observed in process measures and measures that are easily impacted with a single medical visit, such as mammography and cervical cancer screenings. Additionally, the results of this study suggest that PBB incentives can have long-lasting impact on performance improvement, with quality performance either sustained or continually improved over a five-year period for most measures. Empirical results also suggest that, for some measures, the performance improvement response to PBB incentives may be delayed. This is particularly true of measures pertaining to diabetes management.

Table 29

Summary of Research Question 2 Findings Categorized by Performance Metric

Hypothesis	Significant improvement[a] Postperiod 1	Significant improvement[a] Postperiod 2	Performance change Period 1 to Period 2	Hypothesis 2 supported
Research Question 2: Are quality improvements tied to performance-based budgeting sustained over time in U.S. Army hospitals?				
Hypothesis 2: Quality improvements attributable to PBB programs in PBB facilities will be sustained throughout the postperiod after each metric is incentivized.				
Measures				
Population health				
Mammography	Yes	Yes	Improve	Yes
Cervical cancer screening	Yes	Yes	Improve	Yes
Diabetes-A1C screening	No	Yes	Improve	Yes
Diabetes-A1C control	No	Yes	Improve	Yes
Diabetes-LDL control	No	Yes	Improve	Yes
Asthma care	No	No	Improve[b]	No

Note. PBB = performance-based budgeting; LDL = low density lipoproteins; URI = upper respiratory infection; CLABSI = central line-associated bloodstream infection.

[a]Relative to comparison facilities.

[b]PBB had a negative (but not significant) effect on asthma care in both periods, but the negative association in the second period was smaller in absolute value when compared to the first period.

Several sensitivity analyses were conducted in order to determine the robustness of results in light of changes in model specification and exclusion criteria for facilities. One set of sensitivity analyses assessed the relative impact of changes in the specification of the age and time variables. Another sensitivity analysis was conducted to determine the impact of dropping all model covariates. None of these changes produced significant changes to the observed results. Two additional sensitivity analyses altered the exclusion criteria for facilities included in the original model. One analysis excluded all facilities operating in MSMs in order to account for the potential of spillover effects. This exclusion produced slightly larger effect sizes for PBB, indicating that some spillover effects may be present in MSMs.

For example, for many measures the comparison group performed better than the PBB group in the pre-intervention period. In MSMs, patients may have circulated between PBB and non-PBB facilities, which would make the performance changes for those patients less distinguishable in Army facilities in those markets. The fact that removing MSM facilities from the analysis increased PBB effect sizes strengthens the evidence for this inference. Despite this, the relative changes in effect sizes were very small and did not impact the overall interpretation of the results. Another sensitivity analysis assessed the relative impact of changing the thresholds for the minimum number of eligible patients required for inclusion in each analysis. The original model excluded facilities with a denominator of fewer than 5 eligible patients. The sensitivity analyses assessed the relative significance of changing this threshold to a minimum of 10 and 20 eligible patients. This did not have a significant impact on the results.

Another sensitivity analysis evaluated the potential for anticipatory effects on performance in the 6-month period prior to the implementation of PBB incentives. The underlying assumption for this analysis was that organizations might start performance improvement efforts ahead of time in order to maximize the potential for receiving incentives once they are offered. For the majority of performance indicators, there was little evidence of anticipatory effects. However, two indicators (well child visits and low back pain) did exhibit some small performance improvements in the 6-month period prior to the implementation of incentives, suggesting anticipatory effects.

Study Limitations

The results of this study indicate that PBB may be a promising approach for encouraging quality performance improvement in military healthcare facilities. However, this study does have five noteworthy limitations.

The first limitation is that this study divides the intervention and comparison groups along military service lines, with the intervention group including Army facilities and the comparison group including Navy and Air Force facilities. This assignment strategy presents the possibility that unobserved or unmeasured differences between the two groups may result from service-specific differences rather than the effects of PBB. For example, the intervention group (Army) had significantly lower performance on many of the performance metrics in the pre-intervention period in comparison to the Navy and Air Force facilities. This suggests that there may have been factors that differed between Army and comparison facilities that could not be controlled but may be relevant to the different performance trajectories that were measured. Since the two groups were not equal prior to the intervention, the Navy and Air Force facilities may not have been a perfect counterfactual for what may have occurred in Army facilities if PBB had not been implemented. Despite this, there are several research design features that were implemented that help to mitigate this concern.

The use of a fixed-effect model helps to address some concerns about service-specific variations that may confound the interpretation of results. In theory, any potential service-specific time-invariant factor is controlled through the use of *fixed effects* in that each study facility is compared to itself over time with respect to each quality measure. Additionally, all measures were examined for performance trends between the two groups prior to intervention. Though there were some pre-intervention differences in the *levels* of performance, the performance *trends* appeared to be similar for almost all measures. Additionally, for the majority of the measures, there was an observable change in performance that occurred in conjunction with the timing of the incentives for the intervention group. Because this occurred for several measures over different periods of time, it strengthens the inference that the PBB incentives played a key role in these performance changes.

Another limitation is that this study focuses on a narrow set of quality metrics. This was largely due to the fact that the MHS did not collect quality performance data in an integrated manner in the earliest years of PBB. During the beginning stages of the Army's PBB program, each military service collected and reported most of their data independently. Very few measures were included in the early PBB program that had consistent data collection across the three military services. HEDIS measures were a notable exception to this because all HEDIS measures are collected using standardized NCQA protocols, and all three military services participated in collecting these data. As a result, the majority of the quality measures in this study pertain to population health and the use of evidence-based medicine, which have been the focus of the HEDIS program.

It is possible that the results observed in the limited set of quality metrics examined in this study may not produce similar results in other types of measures. As discussed previously, the results of pay for performance studies vary widely depending on the type of quality metric examined (Damberg et al., 2014; Eijkenaar et al., 2013; Van Herck et al., 2010). Though this study includes a mixture of process and outcome measures, most of these metrics are related to the technical aspects of ambulatory care services and do not measure patient or physician perception of quality. According to a policy review by Hanefeld, Powell-Jackson, and Balabanova (2017), comprehensive measures of clinical quality should include aspects of care that go beyond technical quality, including a comparison of clinical quality and perceived quality; measures of quality at varying points in the patients' pathway through the system; measures of immediate and upstream drivers of quality care; and individual and collective assessments of contextual variables such as power, social status, trust and values that impact perception of quality. Though many of these aspects of quality measurement were addressed in measures that the Army included in its PBB programs, not all of these metrics were evaluated in this study due to the consistency issues previously discussed.

A third limitation is that some of the indicators exhibited trends that looked different in the pre-PBB period across the intervention and comparison group (specifically, the URI and CLABSI measures). The findings for these indicators should be interpreted with more caution when compared to the conclusions reached for indicators where the pre-PBB trends appeared to be more similar and stable across the PBB and non-PBB facilities. A critical aspect of the quasi-experimental DID approach is that the intervention and comparison groups must have similar pre-intervention trends in the measurement of the variable in question. Since these pre-intervention *parallel trends* were not observed in the URI and CLABSI measures, there is less strength in the conclusions that can be drawn from these results. In the case of the URI measure, the pre-intervention trends for the comparison group appear to have risen and then levelled off around the time the incentives started. For the intervention group of PBB facilities, the opposite was true. Performance was mostly level until just prior to the start of incentives, and then it rose sharply and continued to rise throughout the majority of the postintervention period. Even though the timing of sharp performance improvement in conjunction with the timing of the incentives suggests that PBB incentives had an impact on performance, this cannot be reliably concluded without similarity in pre-intervention performance trends for the comparison group. With respect to the CLABSI measure, there was not enough consistency in performance from either group in either time period to adequately assess pre- or postintervention performance trends. As mentioned previously, this is likely due to the small sample size. A larger sample may have generated more distinct results and enabled a better observation of pre- and postintervention performance trends between the two groups.

A fourth limitation in this study is the use of suboptimal control variables. The variables for patient characteristics, namely age, gender, and acuity, were derived from the encounters that occurred within the facility, whereas the performance data were derived from the enrolled patient population at each facility. This is due to the fact that the control variables were constructed from data retrieved from the MHS MDR. This data repository stores multiple datapoints for every medical encounter in the MHS. Thus, the patient variables that were aggregated for each facility during each month or quarter were generated from the data on the *visits that occurred* within each facility, not the characteristics of the underlying patient population *enrolled* to that facility. In theory, the medical encounters occurring in each facility should provide a rough approximation of the patient characteristics of all enrollees, though it is possible that these estimates are biased. Thus, the patient characteristics used to construct the control variables may not have been a precise reflection of characteristics of the specific patient population to which each performance measure pertained.

For example, if a diabetes patient is assigned to facility *j* at time *t*, but did not have an appointment at facility *j* during time *t,* then that patient's characteristics (age, gender, acuity) would not have been captured in the data used to generate facility *j*'s aggregated age, gender, and acuity variables. However, that same patient *would* have been included in the facility's diabetes performance measure if the patient was enrolled to facility *j*, regardless of whether the patient had a medical encounter at that facility during the relevant time period. This is because the HEDIS performance measures are constructed from enrollment data, whereas the aggregated control variables are constructed from encounter data. Another related issue exists with the measurement of facility-aggregated patient acuity. Since this measure is derived from patient encounter data, it is likely that acuity estimates are biased by the exclusion of low acuity patients who are assigned to facilities but did not seek medical care. For young, healthy active duty servicemembers, medical encounters may be as infrequent as once per year, when the servicemember receives a required physical exam. This can potentially create a stronger bias in the acuity estimates for those facilities that exclusively serve active duty servicemembers.

The use of suboptimal data in constructing the facility aggregated control variables may have contributed to the lack of significance observed in the estimates of many of these variables. The opposite is true, however, of the DHP (munificence) control variable. The DHP variable was problematic because it was a blunt instrument for approximating the level of munificence in each of the facilities in the study. The DHP budget applied to the entire MHS, meaning that it was time-variant but not facility-variant. Thus, this measure did not capture the distinctions in the availability of resources that may have existed between facilities. Nonetheless, the DHP variable was significant in almost all regression models, although its association with performance was not always in the predicted direction. For some indicators, the size of the DHP budget and performance were inversely related. The reasons for this unexpected association are unclear, but may be correlated with time. Due to concerns about the precision of control variables, a sensitivity analysis was conducted to evaluate the impact of dropping all control variables in the model. This change in model specification did not significantly alter the results or change the interpretation of the findings. Even though the control variables were suboptimal, they did not appear to have a substantial impact on the findings related to potential PBB influence on the quality metrics studied.

A final limitation pertains to the generalizability of the empirical findings. The results of this study suggest that PBB can be an effective tool for encouraging healthcare organizations to improve performance on a select set of quality indicators. However, it is unclear whether these results will generalize to other types of programs or government-funded healthcare organizations. Military healthcare facilities are unique organizations for several reasons that have previously been discussed. Most notably, military healthcare organizations have a very centralized, top-down governance structure that may enable programs such as PBAM and IRIS to generate performance effects that may not be actualized in other types of organizations and governance structures. For example, it is possible that PBB is best situated for performance change in organizations with a uniform set of priorities from a single executive agency (such as the DHA) and may not work as well in other government-funded hospitals that must still contend with multiple priorities from several health insurers. This study did not address the organizational factors that may have contributed to the success of PBB in military healthcare facilities, so it is unknown whether these results will generalize to other types of healthcare organizations.

Implications and Future Research Directions

Theoretical Implications

Most of the literature on PBB has previously been derived from the political science research and has not included rigorous comparative analyses of organizations using and not using PBB. The political science literature on PBB is centered on debates about the most effective mechanism for linking budgetary resources ("inputs") to government "outputs" such as high-performing public programs. Most of this literature is devoted to defining and measuring performance and determining how to most effectively integrate it into the budgeting process. This study offers a new direction for PBB research by examining PBB from an organizational rather than political perspective.

This study draws from the widely used RDT to explain how military healthcare organizations might respond to performance-contingent changes in funding. This theoretical approach was selected over the theories in the political science domain because it focuses on behavior at the organizational level of a healthcare facility as opposed to examining it from the larger governmental and legislative perspective.

The hypotheses generated from the RDT framework and conceptual model were supported for the majority of the indicators in this study. However, this conceptual model may not provide a complete picture of how organizations respond to PBB incentives. It does not explain why PBB had a positive effect for some performance indicators and not others. It also does not explain why PBB was relatively successful in the context of military healthcare organizations, whereas similar approaches, such as pay for performance, have not been as successful in private sector health organizations.

For future analysis, it may be helpful to draw from a wider range of organizational theories to develop a more complete understanding of the factors that may influence the relative success of PBB programs in improving quality performance. For example, future studies might draw upon the elements of contingency theory to determine which types of organizations are most likely to respond positively to performance-based contingencies in funding, such as PBB programs. Future studies might also draw upon the conceptual model posited by William Ocasio's (1997) attention-based view of the firm. Using this framework, PBB might be considered a mechanism by which decision-makers use performance-contingent funding to focus the attention of organizational leaders on the quality issues that are most important to leaders and the organization as a whole. This framework also opens up many opportunities for future research in determining the organizational contexts in which this approach might be most successful.

Policy Implications

This study complements the existing literature on PBB by providing empirical evidence that suggests it can be effective in military healthcare organizations in promoting quality in certain healthcare metrics. It also adds to the wide body of extant literature on pay for performance by examining the specific case of PBB and the unique context of federally funded military healthcare organizations. Additionally, the findings of this study diverge slightly from the trends observed in pay for performance research, which are often contradictory and mixed in terms of offering evidence to support its use. The positive results from this study provide evidence for a type of pay for performance that might be effective in the context of government funded healthcare facilities and programs. This warrants further investigation and replication to determine the contextual factors that may be contributing to the success of PBB in healthcare when much less success has typically been observed for pay for performance.

This study builds upon previous research by West, Cronk, Goodman, and Waymire (2010) that found that the Army's PBAM program demonstrated significant initial improvements in productivity and the use of evidence-based medicine. That study was used to evaluate PBAM and justify its future use. This study extends that line of research by investigating the impact of PBAM and its successor, IRIS, to determine if those programs had ongoing positive impacts on quality performance. This study confirms those initial findings and provides evidence that those initial performance gains were sustained. It also provides evidence that this approach facilitated continued success for other indicators as the programs were expanded to include a wider range of quality measures.

As the MHS continues its integration under the DHA, leaders will have to decide between a range of policy approaches that have been implemented individually across the three service branches. The choice to expand the Army's IRIS program to include all military healthcare facilities is one such decision. This study provides some empirical support to justify its expansion and future use. Though this study did not find evidence that PBB demonstrated positive performance effects for all quality indicators, there were enough instances of success that it should be considered for future use in the areas for which it had significant effects. Because this study did not find PBB to be universally effective across all indicators, it underscores the need for leaders to continually monitor its effects and adjust strategies and incentives accordingly

Finally, the results of this study offer evidence to support the future investigation of PBB as an approach that can potentially be expanded to other government-funded healthcare facilities, such the VHA or the Indian Health Service. It is possible that the success observed in military healthcare facilities may generalize to other types of government-funded healthcare programs.

Suggestions for Future Research

As previously discussed, not all of the indicators demonstrated performance improvement in response to PBB incentives. The well child visits and mental health measures should be examined closely for potential resource shortfalls, barriers to implementation, or other factors that may help to explain their lack of improvement. It would be helpful from a research and policy perspective to better understand which types of performance indicators might be most responsive to incentives so leaders can strategically craft future programs to ensure that budgetary resources are maximized in the areas in which they are likely to be most helpful. It would also be helpful to better understand what (if any) better alternatives exist for indicators that do not respond well to budget incentives. For example, if poor performance is related to resource barriers, then it is possible that resolving these issues may lead to the same types of positive effects as financial incentives. Examining these types of questions may best be accomplished with qualitative research such as grounded theory methods.

Qualitative research might also help to expose some of the underlying organizational factors that either hinder or facilitate the implementation of effective PBB programs. For organizations that are successful in improving performance, qualitative research might help to identify the specific tools or implementation factors that were most advantageous.

Additionally, the results of this study open up many possibilities for future research in that they offer preliminary support for the use of PBB in a military healthcare context but leave many research questions unanswered An important first step is to continue this line of research to confirm that the results replicate in non-Army healthcare facilities. As military hospitals of all three service branches integrate under the management of the DHA and if PBB is expanded to all military facilities, it will be possible to assess the impacts of PBB on a wider selection of military facilities. This can be accomplished using a time-phased comparison of performance change as hospitals integrate into a unified PBB program. A study of this nature will help to address one of the key limitations in this study, namely whether unmeasured service-specific factors may have confounded the interpretation of results. A study of this nature would also offer an opportunity to evaluate a more comprehensive set of quality indicators, as the three services are presently measuring more quality indicators in a uniform manner than during the early stages of PBB adoption.

Future studies might also consider the impact of removing incentives once a high level of performance is achieved. Numerous studies in the pay for performance domain have found evidence of ceiling effects or diminishing performance gains over time, usually once an organization reaches a high level of performance (Ryan, Blustein, & Casalino, 2012; Van Herck et al., 2010; R. Werner et al., 2011). This raises policy questions regarding the point at which incentives could be removed or shifted to other performance areas where they may be more beneficial. Future studies might address this question by studying the impact on performance when incentives are removed after a long period of sustained improvement. It is possible that performance may be sustained because organizations have already developed permanent systems and processes for high performance. It is also possible that the removal of incentives might cause organizations to shift attention to incentivized measures, and performance will decline for measures that are no longer incentivized. Studies that enable more precise estimates of the performance impacts of incentives and their removal will aid decision-makers in crafting policies that maximize the use of performance-based funding incentives.

Another opportunity for future research may be to evaluate the impact of PBB on other types of indicators within the healthcare domain. For example, the PBAM and IRIS programs include productivity and access measures, which have not yet been rigorously studied. Additionally, nontechnical aspects of healthcare such as patient satisfaction, employee satisfaction and well-being, and healthcare equity are also recognized as important indicators for robust medical systems (Hanefeld et al., 2017) and could be included in future research involving PBB.

Conclusion

This study applied RDT as a theoretical model explaining how PBB might impact facility-level performance on various healthcare quality indicators. Research Question 1 drew upon a wide body of extant research on pay for performance to question the impact of a related approach, performance-based budgeting, on quality improvement in U.S. Army healthcare facilities. To address this question, a quasi-experimental, DID analysis was undertaken to compare changes in quality performance among military healthcare facilities that adopted PBB in comparison to healthcare facilities that did not adopt PBB. Performance changes were compared for both groups before and after the implementation of incentives for a variety of quality indicators. The hypothesis was that PBB facilities would demonstrate greater performance improvement on incentivized quality metrics in comparison to military healthcare organizations in other branches that did not participate in PBB programs. This hypothesis was supported for seven out of 13 of the indicators examined, suggesting that PBB can be a helpful tool for encouraging performance improvement on a select group of quality measures in military healthcare organizations.

Research Question 2 was an extension of Research Question 1, asking whether initial performance changes attributable to PBB would be sustained over time. To address this question, six performance measures were examined over a 5-year postperiod. The 5-year period was divided into two 2.5-year periods, and performance was compared between the two periods. It was hypothesized that quality improvements attributable to PBB programs will be sustained in PBB facilities throughout the 5-year postperiod. Supporting Hypothesis 2, performance improved in the second half of the postperiod, relative to the first half, for five of six performance indicators. The results of this portion of the study suggest that once performance incentives are offered, facilities sustain their performance long term or continue to improve.

Several sensitivity analyses were undertaken to examine the robustness of the empirical results. In general, the results for Research Question 1 were robust across all model specifications. A notable finding in one of the sensitivity analyses is that very few of the indicators showed any evidence of significant anticipatory effects. It did not appear that healthcare facilities improved performance prior to the start of the program in anticipation of future incentive opportunities. This finding, taken in conjunction with the some of the findings from Research Question 2, suggests that performance improvement is often sluggish or even delayed in some cases. This appears to be especially true of performance indicators that likely involve a more complex interaction between provider processes and patient variables, such as with the diabetes management measures.

Although this study has several limitations, the results are promising from both practical and theoretical perspectives. The study findings are relevant to clinicians and administrators in military and government-funded healthcare organizations, as they offer evidence to support the future use of PBB as a mechanism for improving quality performance. This study also extends pay for performance research by proffering an argument for classifying PBB as a unique variant of pay for performance and offering evidence that it is effective under certain conditions. As military healthcare leaders continue to grapple with resource allocation decisions to support the highest level of quality healthcare, more studies are needed to examine the impact of resourcing policies such as PBB. This study offers an important first step in understanding the degree to which military healthcare facilities alter performance in response to funding incentives.

References

32 C.F.R. §199.17(p)(5) - TRICARE program. (n.d.). Retrieved from https://www.govregs.com/regulations/expand/title32_chapterI_part199_section199.17

Adirim, T. (2019). A military health system for the twenty-first century. *Health Affairs*, *38*(8), 1268–1273. https://doi.org/10.1377/hlthaff.2019.00302

Agarwal, P. (2019). *Allocative efficiency*. Retrieved November 18, 2019, from https://www.intelligenteconomist.com/allocative-efficiency/

Agency for Healthcare Research and Quality. (2015). *Types of healthcare quality measures*. Retrieved November 8, 2019, from https://www.ahrq.gov/talkingquality/measures/types.html

Allen, T., Mason, T., & Whittaker, W. (2014). Impacts of pay for performance on the quality of primary care. *Risk Management and Healthcare Policy*, *7*, 113–120. https://doi.org/10.2147/RMHP.S46423

American College of Surgeons. (2019). *ACS national surgical quality improvement program*. Retrieved from https://www.facs.org/quality-programs/acs-nsqip

Angrist, J. D., & Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton, NJ: Princeton University Press.

Angrist, J. D., & Pischke, J.-S. (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *SSRN*, *24*(2), 3–30. https://doi.org/10.2139/ssrn.1639809

Balakrishnan, R., Soderstrom, N. S., & West, T. D. (2007). Spending patterns with lapsing budgets: Evidence from U.S. Army hospitals. *Journal of Management Accounting Research*, *19*(1), 1–23. https://doi.org/10.2308/jmar.2007.19.1.1

Banaszak-Holl, J., Zinn, J. S., & Mor, V. (1996). The impact of market and organizational characteristics on nursing care facility service innovation: A resource dependency perspective. *Health Services Research*, *31*(1), 97–117. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8617612%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1070105

Bastian, N. D., Kang, H., Nembhard, H. B., Bloschichak, A., & Griffin, P. M. (2016). The impact of a pay-for-performance program on central line-associated blood stream infections in Pennsylvania. *Hospital Topics*, *94*(1), 8–14. https://doi.org/10.1080/00185868.2015.1130542

Basu, S. (2012). *DoD Moves toward more joint medical services while avoiding a unified command*. Washington, DC: U.S. Medicine. Retrieved from http://www.usmedicine.com/agencies/department-of-defense-dod/dod-moves-toward-more-joint-medical-services-while-avoiding-a-unified-command/

Bazinsky, K., & Bailit, M. (2013, September 10). *The significant lack of alignment across state and regional health measure sets*. Retrieved from http://www.bailit-health.com/articles/091113_bhp_measuresbrief.pdf

Benabbas, R., Shan, G., Akindutire, O., Mehta, N., & Sinert, R. (2019). The effect of pay-for-performance compensation model implementation on vaccination rate: A systematic review. *Quality Management in Health Care*, *28*(3), 155–162. https://doi.org/10.1097/QMH.0000000000000219

Ben-Shalom, Y., Schone, E., & Bannick, R. (2019). Provider acceptance and beneficiary access under TRICARE's PPO health plan. *Health Affairs*, *38*(8), 1343–1350. https://doi.org/10.1377/hlthaff.2019.00220

Benzer, J. K., Young, G. J., Burgess, J. F., Baker, E., Mohr, D. C., Charns, M. P., & Kaboli, P. J. (2014). Sustainability of quality improvement following removal of pay-for-performance incentives. *Journal of General Internal Medicine*, *29*(1), 127–132. https://doi.org/10.1007/s11606-013-2572-4

Berwick, D. M., Nolan, T. W., & Whittington, J. (2008). The triple aim: Care, health, and cost. *Health Affairs*, *27*(3), 759–769. https://doi.org/10.1377/hlthaff.27.3.759

Bolon, D. S. (2005). Comparing mission statement content in for-profit and not-for-profit hospitals: Does mission really matter? *Hospital Topics*, *83*(4), 2–9. https://doi.org/10.3200/HTPS.83.4.2-10

Bond, A. M. (2018). *Nonlinear physician performance contracts* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (10809951)

Bond, A. M., & Schwab, S. D. (2019). Utilization variation in military versus civilian care: Evidence From TRICARE. *Health Affairs*, *38*(8), 1327–1334. https://doi.org/10.1377/hlthaff.2019.00298

Bonfrer, I., Soeters, R., Van de Poel, E., Basenya, O., Longin, G., van de Looij, F., & van Doorslaer, E. (2014). Introduction of performance-based financing in Burundi was associated with improvements in care and quality. *Health Affairs*, *33*(12), 2179–2187. https://doi.org/10.1377/hlthaff.2014.0081

Bono, R. (2017). A more integrated health system: Military medicine and the Defense Health Agency in 2018. *Brief by Director of the Defense Health Agency* [PowerPoint slides], 1–24.

Boukus, E., Cassil, A., & O'Malley, A. S. (2009). A snapshot of U.S. physicians: Key findings from the 2008 Health Tracking Physician Survey. *Data Bull (Cent Stud Health Syst Change), Sept*(35), 1-11.

Burwell, S. M. (2013). *Impacts and costs of the governement shutdown*. Retrieved August 7, 2018, from https://obamawhitehouse.archives.gov/blog/2013/11/07/impacts-and-costs-government-shutdown

Cain, R. (2013, September 4). *Defense Health Agency leans forward to change*. Retrieved from https://www.army.mil/article/110491/Defense_Health_Agency_leans_forward_to_change/

Calikoglu, S., Murray, R., & Feeney, D. (2012). Hospital pay-for-performance programs in Maryland produced strong results, including reduced hospital-acquired conditions. *Health Affairs*, *31*(12), 2649–2658. https://doi.org/10.1377/hlthaff.2012.0357

Carabajal, S. (2012, January). Community based medical homes provide quality care. *Army Medicine*. Retrieved from https://www.army.mil/article/71726/community_based_medical_homes_provide_quality_care

Cassel, C. L., Conway, P. H., Delbanco, S. F., Jha, A. K., Saunders, R. S., & Lee, T. H. (2012). Getting more performance from performance measurement. *New England Journal of Medicine*, 1–4.

Centers for Disease Control and Prevention. (2019). *About NHSN*. Retrieved from https://www.cdc.gov/nhsn/about-nhsn/index.html

Centers for Disease Control and Prevention. (2020). Chapter 4: Bloodstream Infection Event (Central Line-Associated Bloodstream Infection and Non-Central Line-Associated Bloodstream Infection). In *National Healthcare Safety Network (NHSN) Patient Safety Component Manual* (pp. 1-49). Retrieved from https://www.cdc.gov/nhsn/pdfs/pscmanual/4psc_clabscurrent.pdf

Centers for Medicare & Medicaid Services. (2015). *MACRA*. Retrieved from https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/MACRA-MIPS-and-APMs/MACRA-MIPS-and-APMs.html

Centers for Medicare & Medicaid Services. (2018). *The value modifier (VM) program*. Retrieved from https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/VMP/Value-Modifier-VM-or-PVBM.html

Centers for Medicare & Medicaid Services. (2019a). *Hospital-acquired condition reduction program (HACRP).* Retrieved from https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/HAC-Reduction-Program.html

Centers for Medicare & Medicaid Services. (2019b). *Hospital value based purchasing*. Retrieved from https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Hospital-Value-Based-Purchasing-.html

Centers for Medicare & Medicaid Services. (2019c). *NHE historical and projections*. Baltimore, MD. Retrieved from https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NationalHealthAccountsProjected.html

Centers for Medicare & Medicaid Services. (2020). *Quality payment program overview*. Retrieved April 27, 2020, from https://qpp.cms.gov/mips/overview

Chairman of the Joint Chiefs of Staff, (2015). *Joint concept for health services*. Retrieved from http://www.dtic.mil/doctrine/concepts/joint_concepts/jceo.pdf

Chaix-Couturier, C., Durand-Zaleski, I., Jolly, D., & Durieux, P. (2000). Effects of financial incentives on medical practice: Results from a systematic review of the literature and methodological issues. *International Journal for Quality in Health Care*, *12*(2), 133–142. https://doi.org/10.1093/intqhc/12.2.133

Chang, R. E., Lin, S. P., & Aron, D. C. (2012). A pay-for-performance program in Taiwan improved care for some diabetes patients, but doctors may have excluded sicker ones. *Health Affairs*, *31*(1), 93–102. https://doi.org/10.1377/hlthaff.2010.0402

Chee, T. T., Ryan, A. M., Wasfy, J. H., & Borden, W. B. (2016). Current state of value-based purchasing programs. *Circulation*, *133*(22), 2197–2205. https://doi.org/10.1161/CIRCULATIONAHA.115.010268

Chen, L. M., Epstein, A. M., Orav, E. J., Filice, C. E., Samson, L. W., & Joynt Maddox, K. E. (2017). Association of practice-level social and medical risk with performance in the Medicare Physician Value-Based Payment Modifier Program. *JAMA*, *318*(5), 453–461. https://doi.org/10.1001/jama.2017.9643

Chien, A. T., Wroblewski, K., Damberg, C., Williams, T. R., Yanagihara, D., Yakunina, Y., & Casalino, L. P. (2012). Do physician organizations located in lower socioeconomic status areas score lower on pay-for-performance measures? *Journal of General Internal Medicine*, *27*(5), 548–554. https://doi.org/10.1007/s11606-011-1946-8

Childress, C. (2013). *The impact of war on military hospital performance: A study of organizations' response to an environmental jolt*. Virginia Commonwealth University. Retrieved from http://scholarscompass.vcu.edu/etd/3171/

Christianson, J. B., Leatherman, S., & Sutherland, K. (2008). Lessons from evaluations of purchaser pay-for-performance programs: A review of the evidence. *Medical Care Research And Review*, *65*(6 (Supplement)).

Coleman, E. A., Parry, C., Chalmers, S., & Min, S. (2006). The care transitions intervention: Results of a randomized controlled trial. *Comprehensive Care Coordination for Chronically Ill Adults*, *166*, 261–275. https://doi.org/10.1002/9781118785775.ch13

Collins, S. (2015). *Defense Health Agency achieves full operating capability*. Retrieved October 21, 2017, from https://www.defense.gov/News/Article/article/621722/defense-health-agency-achieves-full-operating-capability/

Coplin, W. D., Merget, A. E., & Bourdeaux, C. (2002). The professional researcher as change agent in the government-performance movement. *Public Administration Review*, *62*(6), 699–711. https://doi.org/10.1111/1540-6210.00252

Damberg, C. L., Sorbero, M. E., Lovejoy, S. L., Martsolf, G. R., Raaen, L., & Mandel, D. (2014). Measuring success in health care value-based purchasing programs: Findings from an environmental scan, literature review, and expert panel discussions. *Rand Health Quarterly*, *4*(3), 9. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5161317/

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, *125*(6), 627–668. https://doi.org/10.1037/0033-2909.125.6.627

Defense Health Agency, Wounded Warrior Care Center. (n.d.). *Walter Reed Bethesda National Medical Military Center: Achieving excellence in casualty care*. Retrieved July 11, 2019, from https://www.wwcc.capmed.mil/SitePages/WarriorCare.aspx

Defense Health Agency. (2013). *MHS quadruple aim*. Retrieved February 8, 2019, from https://health.mil/Reference-Center/Glossary-Terms/2013/04/09/MHS-Quadruple-Aim

Defense Health Agency. (2017). *2017 Stakeholder Report*. Retrieved from https://health.mil/Reference-Center/Reports/2018/05/01/Defense-Health-Agency-2017-Stakeholder-Report

Defense Health Agency. (2018a). Evaluation of the TRICARE Program: Fiscal Year 2018 Report to Congress. *Report to Congress*, 1–206.

Defense Health Agency. (2018b). *TRICARE for life*. Retrieved from https://tricare.mil/tfl

Defense Health Agency. (2019a). *About the Military Health System*. Retrieved from https://www.health.mil/About-MHS

Defense Health Agency. (2019b). *Carepoint Information Portal*. Retrieved April 4, 2019, from https://carepoint.health.mil/

Defense Health Agency. (2019c). *Defense Health Agency*. Retrieved February 14, 2019, from https://www.health.mil/About-MHS/OASDHA/Defense-Health-Agency

Defense Health Agency. (2019d). *Elements of the MHS*. Retrieved from https://health.mil/About-MHS/MHS-Elements

Defense Health Agency. (2019e). *Evaluation of the TRICARE program: Fiscal year 2019 report to Congress*. Retrieved from http://health.mil/Reference-Center/Reports/2016/05/19/Evaluation-of-the-TRICARE-Program-Fiscal-Year-2016-Report-to-Congress

Defense Health Agency. (2019f). *Military Health System Data Repository*. Retrieved March 4, 2019, from https://www.health.mil/

Defense Health Agency. (2019g). *Military hospitals and clinics*. Retrieved July 7, 2019, from https://www.tricare.mil/FindDoctor/AllProviderDirectories/Military.aspx

Defense Health Agency. (2019h). *Multi-service markets*. Retrieved February 10, 2019, from https://www.health.mil/Military-Health-Topics/Access-Cost-Quality-and-Safety/Access-to-Healthcare/Multi-Service-Markets

Defense Health Agency. (2020). *Defense medicial information system identification tables*. Falls Church, Virginia: health.mil. Retrieved from https://www.health.mil/Military-Health-Topics/Technology/Support-Areas/Geographic-Reference-Information/DMIS-ID-Tables

DeVries, A., Li, C.-H. W., Sridhar, G., Hummel, J. R., Breidbart, S., & Barron, J. J. (2012). Impact of medical homes on quality, healthcare utilization, and costs. *The American Journal of Managed Care*, *18*(9), 534–544. Retrieved from http://europepmc.org/abstract/MED/23009304

DeVries, M. S., & Nemec, J. (2019). Dilemmas in performance-based budgeting. In M. S. DeVries & J. Nemec (Eds.), *Performance-based budgeting in the public sector* (pp. 3–25). Retrieved from http://ebookcentral.proquest.com

Dietrich, E. J. (2013). *Pay for performance in the public sector: From the National Security Personnel System to the Affordable Care Act*. American University. https://doi.org/10.1192/bjp.112.483.211-a

Dolfini-Reed, M., & Jebo, J. (2000). The evolution of the military health care system: Changes in public law and DOD regulations. *Center For Naval Analysis*, *1*(July). Retrieved from http://www.cna.org/sites/default/files/research/d0000437.a3.pdf

Donabedian, A. (1980). *Explorations in quality assessment and monitoring*. Ann Arbor, MI: Health Administration Press.

Doran, T. (2014). Incentivising improvements in health care delivery: A response to Adam Oliver. *Health Economics, Policy and Law*, *10*(3), 351–356. https://doi.org/10.1017/S1744133114000498

Doran, T., Fullwood, C., Gravelle, H., Reeves, D., Kontopantelis, E., Hiroeh, U., & Roland, M. (2006). Pay-for-performance programs in family practices in the United Kingdom. *New England Journal of Medicine*, *355*(17), 1832–1833. https://doi.org/10.1056/NEJMc062305

Doran, T., Kontopantelis, E., & Valderas, J. M. (2011). Effect of financial incentives on incentivised and non-incentivized clinical activities: Longitudinal analysis of data from the UK quality and outcomes framework. *BMJ*, *342*, 1–12. https://doi.org/10.1136/bmj.d3590

Doran, T., Maurer, K. A., & Ryan, A. (2017). Impact of provider incentives on quality and value of health care. *Annual Review of Public Health*, *38*, 449–465. https://doi.org/10.1146/annurev-publhealth-032315-021457

Dunning, D. K. (2014). Integrated resourcing and incentive system. *Internal Brief by United States Medical Command Assistant Chief of Staff* [PowerPoint slides], 1–25.

Eijkenaar, F., Emmert, M., Scheppach, M., & Schoffski, O. (2013). Effects of pay for performance in health care: A systematic review of systematic reviews. *Health Policy*, *110*(2–3), 115–130. https://doi.org/10.1016/j.healthpol.2013.01.008

Elixhauser, A., Steiner, C., & Harris, D. R. (1998). Comorbidity measures for use with administrative data. *Medical Care*, *36*(1), 8–27.

Emmert, M., Eijkenaar, F., Kemter, H., Esslinger, A. S., & Schoffski, O. (2012). Economic evaluation of pay-for-performance in health care: A systematic review. *European Journal of Health Economics*, *13*(6), 755–767. https://doi.org/10.1007/s10198-011-0329-8

Encinosa, W. E., & Hellinger, F. J. (2008). The impact of medical errors on ninety-day costs and outcomes: An examination of surgical patients. *Health Services Research*, *43*(6), 2067–2085. https://doi.org/10.1111/j.1475-6773.2008.00882.x

Fareed, N., & Mick, S. S. (2011). To make or buy patient safety solutions : A resource dependence and transaction. *Health Care Management Review*, *36*(4), 288–298. https://doi.org/10.1097/HMR.0b013e318225998b

Felt-Lisk, S., Gimm, G., & Peterson, S. (2007). Making pay-for-performance work in medicaid. *Health Affairs*, *26*(4), 516–527. https://doi.org/10.1377/hlthaff.26.4.w516

Feng, Y., Kristensen, S. R., Lorgelly, P., Meacock, R., Sanchez, M. R., Siciliani, L., & Sutton, M. (2019). Pay for performance for specialised care in England: Strengths and weaknesses. *Health Policy, 123*(11), 1036–1041.

https://doi.org/10.1016/j.healthpol.2019.07.007

Figueroa, J. F., Tsugawa, Y., Jie Zheng, E., Orav, J., & Jha, A. K. (2016). Association between the value-based purchasing pay for performance program and patient mortality in U.S. hospitals: Observational study. *BMJ (Online)*, *353*. https://doi.org/10.1136/bmj.i2214

Flodgren, G., Eccles, M. P., Shepperd, S., Scott, A., Parmelli, E., & Beyer, F. R. (2011). An overview of reviews evaluating the effectiveness of financial incentives in changing healthcare professional behaviours and patient outcomes. *Cochrane Database of Systematic Reviews*, *7*. https://doi.org/10.1002/14651858.cd009255

Forster, A. J., Murff, H. J., Peterson, J. F., Gandhi, T. K., & Bates, D. W. (2003). The incidence and severity of adverse events affecting patients after discharge from the hospital. *Annals of Internal Medicine*, *138*, 161–174.

Frakes, M., & Gruber, J. (2019). Defensive medicine: Evidence from military immunity. *American Economic Journal: Economic Policy*, *11*(3), 197–231. https://doi.org/10.1257/pol.20180167

Gillam, S. J., Niroshan Siriwardena, A., & Steel, N. (2012). Pay-for-performance in the United Kingdom: Impact of the quality and outcomes framework—A systematic review. *Annals of Family Medicine*, *10*(5), 461–468. https://doi.org/10.1370/afm.1377

Glickman, S. W., Ou, F.-S., DeLong, E. R., Roe, M. T., Lytle, B. L., Mulgund, J., . . . Peterson, E. D. (2007). Pay for performance, quality of care, and outcomes in acute myocardial infarction. *JAMA*, *297*(21), 2373–2380. https://doi.org/10.1001/jama.297.21.2373

Gray, A. M., Phillips, V. L., & Normand, C. (1996). The costs of nursing turnover: Evidence from the British National Health Service. *Health Policy*, *38*(2), 117–128. https://doi.org/10.1016/0168-8510(96)00854-8

Green, J., Passman, L. J., & Wintfield, N. (1991). Analyzing hospital mortality. *JAMA*, *265*(14), 1849–1853. https://doi.org/10.1001/jama.1991.03460140077030

Griffith, R. (2013). *Performance based adjustment model overview*. Presentation at the 2013 Healthcare Financial Management Association Texas State Conference, Austin, Texas.

Griffith, R. (2019, September 24). *Interview on Army medical department performance-based budgeting programs (2006-2019).* Fort Sam Houston, TX: Joint Base San Antonio.

Gupta, N., & Ayles, H. M. (2019). Effects of pay-for-performance for primary care physicians on diabetes outcomes in single-payer health systems: A systematic review. *The European Journal of Health Economics*, *0123456789*. https://doi.org/10.1007/s10198-019-01097-4

Hanefeld, J., Powell-Jackson, T., & Balabanova, D. (2017). Understanding and measuring quality of care: Dealing with complexity. *Bull World Health Organization*, *95*, 368–374.

https://doi.org/10.3109/03008208109152129

Harman, J. S., Scholle, S. H., Ng, J. H., Pawlson, L.G., Mardon, R. E., Haffer, S.C., Shih, S., Bierman, A. S. (2010). Association of Health Plans' Healthcare Effectiveness Data and Information Set (HEDIS) performance with outcomes of enrollees with diabetes. *Medical Care*, *48*(3), 217–223. https://doi.org/10.1097/MLR.0b013e3181ca3fe6

Hearld, L. R., Alexander, J. A., Fraser, I., & Jiang, H. J. (2008). How do hospital organizational structure and processes affect quality of care? A critical review of research methods. *Medical Care Research and Review*, *65*(3), 259–299. https://doi.org/10.1177/1077558707309613

Herbst, T., & Emmert, M. (2017). Characterization and effectiveness of pay-for-performance in ophthalmology: A systematic review. *BMC Health Services Research*, *17*(1), 1–11. https://doi.org/10.1186/s12913-017-2333-x

Hosek, S. D., & Cecchine, G. (2001). Medical readiness and operational medicine. In *Reorganizing the military health system: Should there be a joint command?* (pp. 43–55). Santa Monica, CA: RAND Corporation. Retrieved from https://www.rand.org/pubs/monograph_reports/MR1350.html.

Houle, S. K., McAlister, F. A., Jackevicius, C. A., Chuck, A. W., & Tsuyuki, R. T. (2012). Does performance-based remuneration for individual health care practitioners affect patient care? A systematic review. *Annals of Internal Medicine*, *157*(12), 889–899. https://doi.org/10.7326/0003-4819-157-12-201212180-00009

Hsieh, H. M., Clement, D. G., & Bazzoli, G. J. (2010). Impacts of market and organizational characteristics on hospital efficiency and uncompensated care. *Health Care Management Review*, *35*(1), 77–87. https://doi.org/10.1097/HMR.0b013e3181c09956

Huang, J., Yin, S., Lin, Y., Jiang, Q., He, Y., & Du, L. (2013). Impact of pay-for-performance on management of diabetes: A systematic review. *Journal of Evidence-Based Medicine*, *6*(3), 173–184. https://doi.org/10.1111/jebm.12052

Huber, P. J. (1967). The behavior of maximum likelihood estimates under non-standard conditions. *Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics* (pp. 221–233). Berkeley: University of California Press. https://projecteuclid.org/euclid.bsmsp/1200512974

Hutter, P. J., Roski, J., Woodson, J., Middleton, A., Kneeland, R., Worthy, A., . . . Cooper, E. (2019). Readiness of medical providers in the military health system: Overview of operational and policy considerations. *Health Affairs*, *38*(8), 1274–1280. https://doi.org/10.1377/hlthaff.2019.00336

Iezzoni, L. (2003). Range of Risk Factors. In L. Iezzoni (Ed.), *Risk Adjustment for Measuring Health Care Outcomes* (3rd ed., pp. 33–70). Chicago: Health Administration Press.

Institute of Medicine. (1993). *Access to health care in America*. Washington, DC: National Academy Press.

Institute of Medicine. (2000). *To err is human: Building a safer health system*. Washington, DC: National Academy Press.

Institute of Medicine. (2001). *Crossing the quality chasm: A new health system for the 21st century*. Washington, DC: National Academy Press. Retrived from https://doi.org/10.17226/10027

Institute of Medicine. (2007). *Rewarding provider performance: Aligning incentives in Medicare* (Pathways to Quality Health Care Series). Washington, DC: National Academies Press. Retrieved from http://books.google.com/books?id=NRPi1qiTWfkC&pgis=1

Institute of Medicine (U.S.), Division of Health Care Services. (1990). *Medicare: A strategy for quality assurance*. (K. N. Lohr, Ed.). Washington, DC: National Academy Press.

James, J. (2012, October 11). Health policy brief: Pay for performance. *Health Affairs Health Policy Brief*, 1-6. https://doi.org/ 10.1377/hpb20121011.90233

Jha, A. K. (2017). Payment power to the patients. *JAMA*, *318*(1), 18–19. https://doi.org/10.1001/jama.2017.7533

Jha, A. K., Joynt, K. E., Orav, E. J., & Epstein, A. M. (2012). The long-term effect of premier pay for performance on patient outcomes. *Obstetrical & Gynecological Survey*, *67*(7), 391–393. https://doi.org/10.1097/01.ogx.0000418567.29397.e2

Joint Chiefs of Staff. (2017). *Joint Concept for Health Services*. Government Printing Office. Retrieved from http://www.dtic.mil/doctrine/concepts/joint_concepts/jceo.pdf

Jones, C. B. (2005). The costs of nurse turnover, Part 2. *The Journal of Nursing Administration*, *35*(1), 41–49. https://doi.org/10.1097/00005110-200501000-00014

Jones, T. L. (2018). Outcome measurement in nursing : Imperatives, ideals, history and challenges. *The Online Journal of Issues in Nursing*, *21*(2), 1–19. https://doi.org/10.3912/OJIN.Vol21No02Man01

Jordan, M. M., & Hackbart, M. M. (1999). Performance budgeting and funding in the states: A status assessment. *Public Budgeting and Finance*, *19*(1), 68–88.

Kane, R. L., & Radosevich, D. M. (2011). *Conducting health outcomes research*. Sudbury, MA: Jones and Bartlett.

Kazley, A. S., & Ozcan, Y. A. (2007). Organizational and environmental determinants of hospital EMR adoption: A national study. *Journal of Medical Systems*, *31*(5), 375–384. https://doi.org/10.1007/s10916-007-9079-7

Kellermann, A. (2017). Rethinking The United States' military health system. *Health Affairs Blog*, *April*, 1–10. https://doi.org/10.1377/hblog20170427.059833

Kime, P. (2014, October 14). Military health system: Patients say they can't get in to see their doctors. *The Military Times [Online]*, pp. 1–6. Retrieved from https://www.militarytimes.com/news/pentagon-congress/2014/10/13/military-health-system-patients-say-they-can-t-get-in-to-see-their-doctors/

Koehlmoos, T. P., Madsen, C. K., Banaag, A., Haider, A. H., Schoenfeld, A. J., & Weissman, J. S. (2019). Assessing Low-Value Health Care Services In The Military Health System. *Health Affairs*, *38*(8), 1351–1357. https://doi.org/10.1377/hlthaff.2019.00252

Kondo, K., Damberg, C., Mendelson, A., Motu'apuaka, M., Freeman, M., O'Neil, M., . . . Kansagara, D. (2015). *Understanding the intervention and implementation factors associated with benefits and harms of pay for performance programs in healthcare*. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/27054229

Kondo, K. K., Wyse, J., Mendelson, A., Beard, G., Freeman, M., Low, A., & Kansagara, D. (2018). Pay-for-performance and veteran care in the VHA and the community: A systematic review. *Journal of General Internal Medicine*, (1), 1–12. https://doi.org/10.1007/s11606-018-4444-4

Kong, D. (2005). Performance-based budgeting: The U.S. experience. *Public Organizational Review: A Global Journal*, *5*, 91–107.

Kouides, R. W., Bennett, N. M., Lewis, B., Cappuccio, J. D., Barker, W. H., & LaForce, F. M. (1998). Performance-based physician reimbursement and influenza immunization rates in the elderly. *American Journal of Preventive Medicine*, *14*(2), 89–95. https://doi.org/10.1016/S0749-3797(97)00028-7

LaFraniere, S., & Lehren, A. (2014a). In military care, a pattern of errors but not scrutiny. *New York Times* (June 28). Retrieved from https://www.nytimes.com/2014/06/29/us/in-military-care-a-pattern-of-errors-but-not-scrutiny.html

LaFraniere, S., & Lehran, A. W. (2014b, September 2). Smaller hospitals aid to put patients at risk: Lack of cases can compromise treatment—Pentagon may cut back system. *New York Times*. Retrieved from https://www.nytimes.com/2014/09/02/us/smaller-military-hospitals-said-to-put-patients-at-risk.html

Lang, T. A., Hodge, M., Olson, V., Romano, P. S., & Kravitz, R. L. (2004). Nurse-patient ratios: A systematic review on the effects of nurse staffing on patient, nurse employee, and hospital outcomes. *JONA: The Journal of Nursing Administration*, *34*(7), 326–337. https://doi.org/10.1097/00005110-200407000-00005

Lauth, T. P. (1985). Performance evaluation in the Georgia budgetary process. *Public Budgeting and Finance*, *5*(1), 67–82. Retrieved from http://onlinelibrary.wiley.com/journal/10.1111/%28ISSN%291540-5850/issues

Lee, J. Y., Lee, S. Il, Kim, N. S., Kim, S. H., Son, W. S., & Jo, M. W. (2012). Healthcare organizations' attitudes toward pay-for-performance in Korea. *Health Policy*, *108*(2–3), 277–285. https://doi.org/10.1016/j.healthpol.2012.09.002

Lester, H., Matharu, T., Mohammed, M. A., Lester, D., & Foskett-Tharby, R. (2013). Implementation of pay for performance in primary care: A qualitative study 8 years after introduction. *British Journal of General Practice*, *63*(611), 408–415. https://doi.org/10.3399/bjgp13X668203

Lin, Y., Yin, S., Huang, J., & Du, L. (2016). Impact of pay for performance on behavior of primary care physicians and patient outcomes. *Journal of Evidence-Based Medicine*, *9*(1), 8–23. https://doi.org/10.1111/jebm.12185

Lindenauer, P. K., Remus, D., Roman, S., Rothberg, M. B., Benjamin, E. M., Ma, A., & Bratzler, D. W. (2007). Public reporting and pay for performance in hospital quality improvement. *New England Journal of Medicine*, *356*, 486–496. Retrieved from http://www.nejm.org/doi/full/10.1056/NEJMsa064964

Linking program funding to performance results: Joint hearing before the subcommitttee on Governement Efficiency, Financial Management and Intergovernemental Relations of the Committeee on Governement Reform. (2002). Washington, DC: U.S. Government Printing Office. Retrieved from http://www.gpo.gov/congress/house

Long, J. S. (1997). *Regression models for categorical and limited dependent variables*. Thousand Oaks, CA: Sage.

Medical and Dental Care in the Armed Forces, 10 USC Ch. 55§§1071–1110b. Retrieved from https://uscode.house.gov/view.xhtml?path=/prelim@title10/subtitleA/part2/chapter55&edition=prelim

Military Compensation and Retirement Modernization Commission. (2015). *Report of the Military Compensation and Retirement Modernization Commission: final report*. Retrieved from https://permanent.access.gpo.gov/gpo54961/MCRMC-FinalReport-29JAN15-HI.pdf

Mangelsdorff, A. D., & Finstuen, K. (2003). Patient satisfaction in military medicine: Status and an empirical test of a model. *Military Medicine*, *168*(9), 744–749. https://doi.org/10.1093/milmed/168.9.744

Markovitz, A. A., & Ryan, A. M. (2017). Pay-for-performance: Disappointing results or masked heterogeneity? *Medical Care Research and Review*, *74*(1), 3–78. https://doi.org/10.1177/1077558715619282

Martin, L., Nelson, E., Lloyd, R., & Nolan, T. (2007). *Whole system measures*. Cambridge, MA:

Institute for Healthcare Improvement.

Mauro, S. G., Cinquini, L., & Grossi, G. (2017). Insights into performance-based budgeting in the public sector: A literature review and a research agenda. *Public Management Review*, *19*(7), 911–931. https://doi.org/10.1080/14719037.2016.1243810

McGill, R. (2001). Performance budgeting. *International Journal of Public Sector Management*, *14*(5), 376–390. https://doi.org/10.1057/9781137315304

McKenzie, D. (2020). Revisiting the Difference-in-Differences Parallel Trends Assumption: Part I Pre-Trend Testing. Retrieved April 6, 2020, from https://blogs.worldbank.org/impactevaluations/revisiting-difference-differences-parallel-trends-assumption-part-i-pre-trend

McNab, R. M., & Melese, F. (2003). Implementing the GPRA: Examining the prospects for performance budgeting in the federal government. *Public Budgeting and Finance*, *23*(2), 73–95. https://doi.org/10.1111/1540-5850.2302006

Menachemi, N., Mazurenko, O., Kazley, A. S., Diana, M. L., & Ford, E. W. (2012). Market factors and electronic medical record adoption in medical practices. *Health Care Management Review*, *37*(1), 14–22. https://doi.org/10.1097/HMR.0b013e3182352562

Menachemi, N., Shin, D. Y., Ford, E. W., & Yu, F. (2011). Environmental factors and health information technology management strategy. *Health Care Management Review*, *36*(3), 275–285. https://doi.org/10.1097/HMR.0b013e3182048e7e

Mendelson, A., Kondo, K., Damberg, C., Low, A., Motuapuaka, M., Freeman, M., . . . Kansagara, D. (2017). The effects of pay-for-performance programs on health, health care use, and processes of care: A systematic review. *Annals of Internal Medicine*, *166*(5), 341–353. https://doi.org/10.7326/M16-1881

Mendez, B. (2018a). *CRS in focus 10530. Defense primer: The military health system*. Washington, DC: Congressional Research Service. Retrieved from https://crsreports.congress.gov

Mendez, B. (2018b). *Military medical care: Frequently asked questions* (CRS Report No. 45399, Version 4). Washington, DC: Congressional Research Service.

Milstein, R., & Schreyoegg, J. (2016). Pay for performance in the inpatient sector: A review of 34 P4P programs in 14 OECD countries. *Health Policy*, *120*(10), 1125–1140. https://doi.org/10.1016/j.healthpol.2016.08.009

Mullen, K. J., Frank, R. G., & Rosenthal, M. B. (2010). Can you get what you pay for? Pay-for-performance and the quality of healthcare providers. *RAND Journal of Economics*, *41*(1), 64–91. https://doi.org/10.1111/j.1756-2171.2009.00090.x

Natarajan, Y., & Kanwal, F. (2015). Pay for performance in chronic liver disease. *Clinical Gastroenterology and Hepatology*, *13*(12), 2042–2047. https://doi.org/10.1016/j.cgh.2015.06.038

National Committee for Quality Assurance. (2019). *HEDIS measures*. Retrieved February 14, 2019, from https://www.ncqa.org/hedis/measures/

National Defense Authorization Act of 2017, Pub. L. No. 114-328, 130 Stat. 2193 (2016). Retrived from https://www.congress.gov/114/plaws/publ328/PLAW-114publ328.pdf

National Health Policy Forum. (2015, January). The basics: Relative value units, 1–5. Retrieved from https://www.nhpf.org/library/the-basics/Basics_RVUs_01-12-15.pdf

National Perinatal Information Center. (2019). *Perinatal center database*. Retrieved March 4, 2019, from http://www.npic.org/

Neumann, P. J., & Levine, B. S. (2002). Do HEDIS measures reflect cost-effective practices? *American Journal of Preventive Medicine*, *23*(4), 276–289. https://doi.org/10.1016/S0749-3797(02)00516-0

NPR.org. (2011, August 31). Walter Reed was the Army's wake-up call in 2007 [Radio transcript]. *Morning Edition*, pp. 1–14. Retrieved from https://www.npr.org/2011/08/31/139641856/in-2007-walter-reed-was-the-armys-wakeup-call

Ocasio, W. (1997). Towards an attention-based view of the firm. *Strategic Management Journal*, *18*(Summer Special Issue), 187–206. https://doi.org/10.1016/S0021-9290(96)00101-7

Office of Management and Budget (OMB). (2011, August 17). OMB Memorandum M-11-31: Delivering an efficient, effective and accountable government [Memorandum]. Washington, DC: Lew, J.J. Retrieved from https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/memoranda/2011/m11-31.pdf

OECD Organization for Economic Cooperation and Development. (2004). Public Sector Modernization: Governing for Performance. *OECD Observer*, (October 2004).

Organization for Economic Cooperation and Development. (2007). *Performance budgeting in OECD countries*. Retrieved from http://proxy.library.vcu.edu/login?url=https://search.proquest.com/docview/58748554?accountid=14780

Pellerin, C. (2017). *Surgeons General emphasize readiness as priority in military medicine*. Washington, DC. Retrieved from https://dod.defense.gov/News/Article/Article/1134247/surgeons-general-emphasize-readiness-as-priority-in-military-medicine/

Performance-based budgeting: Hearing before the Committee on Budget, House of Respresentatives, 109th Cong., first session. (2005). Washington, DC.

Petersen, L. A., Simpson, K., Pietz, K., Urech, T. H., Hysong, S. J., Profit, J., . . . Woodard, L. C. D. (2013). Effects of individual physician-level and practice-level financial incentives on hypertension care: A randomized trial. *JAMA*, *310*(10), 1042–1050. https://doi.org/10.1001/jama.2013.276303

Petrullo, K. A., Lamar, S., Nwankwo-Otti, O., Alexander-Mills, K., & Viola, D. (2012). The Patient Satisfaction Survey: What does it mean to your bottom line? *Journal of Hospital Administration*, *2*(2), 1–8. https://doi.org/10.5430/jha.v2n2p1

Pfeffer, J., & Salancik, G. R. (1978). The design and management of externally controlled organizations. In *The external control of organizations* (pp. 257–287). Stanford, CA: Stanford University Press.

Powell, L. R. (2018). *CLABSI definition with case studies NHSN bloodstream infection surveillance in 2018*. U.S. Centers for Disease Control and Prevention: National Center for Emerging and Zoonotic Infectious Diseases. Retrieved from https://www.cdc.gov/nhsn/pdfs/training/2018/clabsi-508.pdf

Priest, D., & Hull, A. (2007, February 18). Soldiers face neglect, frustration at Army's top medical facility. *The Washington Post*, pp. 1–10. Retrieved from file:///N:/Projects/DODPDH/Library/COMPREHENSIVE_LIBRARY--DO_NOT_ADD-REMOVE/Priest(18Feb2007)WashPost-SldrFaceNglctFrust@ArmysTopMedclFac.pdf

Radin, B. A. (2000). The Government Performance and Results Act and the tradition of federal management reform: Square pegs in round holes? *Journal of Public Administration Research and Theory*, *10*(1), 111–135. https://doi.org/10.1093/oxfordjournals.jpart.a024260

Ramanujan, R., & Rousseau, D. M. (2006). The challenges are organisatinal not just clinical. *Journal of Organisational Behaviour*, *27*(7), 811–827.

Rheney, C. (2007). PBAM & productivity: A primer for performance. *Internal brief by the Chief Financial Offcier of Dewitt Healthcare Network.* [Powerpoint slides, unpublished], 1-32.

Robinson, M., & Brumby, J. (2005). Does performance budgeting work? An analytical review of the empirical literature. *IMF Working Papers*, *05*(210), 1. https://doi.org/10.5089/9781451862294.001

Roland, M., & Campbell, S. (2014). Health policy report: Successes and failures of pay for performance in the United Kingdom. *New England Journal of Medicine*, *370*(20), 1944–1949.

Rosenthal, M. B., & Dudley, R. A. (2007). Will the latest payment trend improve care? *JAMA*, *297*(7), 740–743.

Rosenthal, M. B., Fernandopulle, R., Song, H. S. R., & Landon, B. (2004). Paying for quality: Providers' incentives for quality improvement. *Health Affairs*, *23*(2), 127–141. https://doi.org/10.1377/hlthaff.23.2.127

Rosenthal, M. B., Landon, B. E., Normand, S. L. T., Frank, R. G., & Epstein, A. M. (2006). Pay for performance in commercial HMOs. *New England Journal of Medicine*, *355*(18), 1895–1902. https://doi.org/10.1056/NEJMsa063682

Rosenthal, M. B., Li, Z., Robertson, A. D., & Milstein, A. (2009). Impact of financial incentives for prenatal care on birth outcomes and spending. *Health Services Research*, *44*(5 PART 1), 1465–1479. https://doi.org/10.1111/j.1475-6773.2009.00996.x

Rubin, D. B. (1977). Assignment to treatment group on the basis of a covariate. *Journal of Educational Statistics*, *2*(1), 1–26. https://doi.org/10.2307/1164933

Rude, T., York, N., Donin, N., Angeles, L., Wysock, J., Makarov, D., . . . York, N. (2017). Anaysis of national trends in hospital acquired conditions following major urologic surgery after implementation of the hosptal acquired condition reduction program. *Journal of Urology*, *197*(4), e1300. https://doi.org/10.1016/j.juro.2017.02.3037

Ryan, A. M. (2009). Effects of the premier hospital quality incentive demonstration on Medicare patient mortality and cost: Quality and performance. *Health Services Research*, *44*(3), 821–842. https://doi.org/10.1111/j.1475-6773.2009.00956.x

Ryan, A. M., Blustein, J., & Casalino, L. P. (2012). Medicare's flagship test of pay-for-performance did not spur more rapid quality improvement among low-performing hospitals. *Health Affairs*, *31*(4), 797–805. https://doi.org/10.1377/hlthaff.2011.0626

Ryan, A. M., Blustein, J., Doran, T., D. Michelow, M., & Casalino, L. P. (2012). The effect of phase 2 of the premier hospital quality incentive demonstration on incentive payments to hospitals caring for disadvantaged patients. *Health Services Research*, *47*(4), 1418–1436. https://doi.org/10.1111/j.1475-6773.2012.01393.x

Ryan, A. M., Burgess, J. F., & Dimick, J. B. (2015). Why we should not be indifferent to specification choices for difference-in-differences. *Health Services Research*, *50*(4), 1211–1235. https://doi.org/10.1111/1475-6773.12270

Ryan, A. M., Burgess, J. F., Pesko, M. F., Borden, W. B., & Dimick, J. B. (2015). The early effects of Medicare's mandatory hospital pay-for-performance program. *Health Services Research*, *50*(1), 81–97. https://doi.org/10.1111/1475-6773.12206

Sanchez, E. (2018, January 29). *BAMC Re-verified as Level I Trauma Center*. Retrieved from https://www.army.mil/article/199770/bamc_re_verified_as_level_i_trauma_center

Schick, A. (2003). The performing state: Reflection on an idea whose time has come but whose implementation has not. *OECD Journal on Budgeting*, *3*(2), 71–103. https://doi.org/10.1787/budget-v3-art10-en

Schwab, S. D. (2018). *You had me at hello: The effects of disruptions to the patient-physician relationship* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 2055979662)

Scott, A., Sivey, P., Ait Ouakrim, D., Willenberg, L., Naccarella, L., Furler, J., & Young, D. (2011). The effect of financial incentives on the quality of health care provided by primary care physicians. *Cochrane Database of Systematic Reviews*, (9). https://doi.org/10.1002/14651858.cd008451.pub2

Seshadri, R., Strane, D., Matone, M., Ruedisueli, K., & Rubin, D. M. (2019). Families with TRICARE report lower health care quality and access compared to other insured and uninsured families. *Health Affairs*, *38*(8), 1377–1385. https://doi.org/10.1377/hlthaff.2019.00274

Sherry, T. B. (2016). A note on the comparative statics of pay-for-performance in health care. *Health Economics*, *25*, 637–644. https://doi.org/10.1002/hec

Shih, T., Nicholas, L. H., Thumma, J. R., Birkmeyer, J. D., & Dimick, J. B. (2014). Does pay-for-performance improve surgical outcomes An evaluation of phase 2 of the premier hospital quality incentive demonstration. *Annals of Surgery*, *259*(4), 677–681. https://doi.org/10.1097/SLA.0000000000000425

Smith, D. J., Bono, R. C., & Slinger, B. J. (2017). Transforming the military health system. *JAMA*, *318*(24), 2427–2428. https://doi.org/10.1001/jama.2017.16718

Spence Laschinger, H. K., & Leiter, M. P. (2006). The impact of nursing work environments on patient safety outcomes: The mediating role of burnout/engagement. *Journal of Nursing Administration*, *36*(5), 259–267. https://doi.org/10.1097/00005110-200605000-00019

Stoto, M. A. (2017). Population health measurement: Applying performance measurement concepts in population health settings. *Solving Population Health Problems Through Collaboration*, *2*, 160–172. https://doi.org/10.4324/9781315212708

Tanielian, T., & Farmer, C. (2019). The U.S. military health system: Promoting readiness and providing health care. *Health Affairs*, *38*(8), 1259–1267. https://doi.org/10.1377/hlthaff.2019.00239

The Joint Commission. (2018). *Facts about ORYX® for hospitals (National Hospital Quality Measures)*. Retrieved March 4, 2019, from https://www.jointcommission.org/facts_about_oryx_for_hospitals/

TRICARE. (2019). *Multi-service markets*. Retrieved June 4, 2019, from https://www.tricare.mil/FindDoctor/AllProviderDirectories/MSM

U.S. Army. (2012). *Army medicine facilities: Landstuhl Regional Medical Center*. Retrieved July 11, 2019, from https://www.goarmy.com/amedd/health-care/facilities/landstuhl-regional-medical-center.html

U.S. Army Medical Command. (2016, June 2). *Integrated Resourcing and Incentive System user guide, Version 5*.[Internally published guide].

U.S. Army Medical Command. (2017). *Integrated Resourcing and Incentive System (IRIS) fiscal year 2017 user guide*. .[Internally published guide].

U.S. Army Medical Department. (2016). *Warrior and family support center*. Retrieved July 11, 2019, from https://www.bamc.health.mil/military/wfsc/

U.S. Department of Defense. (2013). *Department of Defense Directive No. 5136.13: Defense Health Agency*. Washington, DC. Retrieved from http://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/513613p.pdf

U.S. Department of Defense. (2011). *Health Affairs Policy 11-005: TRICARE policy for access to care*. Washington, DC: Author.

U.S. Department of Defense. (2014). *Military health system review*. Washington, DC: Author.

U.S. Department of Defense. (2016). *Department of Defense Instruction No.6200.06: Periodic health assessment (PHA) program*. Washington, DC: Author.

U.S. Department of Defense. (2018a). *Defense Health Program. Fiscal Year (FY) 2019 Budget Estimates*.

U.S. Department of Defense. (2018b). *Demographics 2018: Profile of the military community*. Retrieved from https://www.militaryonesource.mil/reports-and-surveys/infographics/active-duty-member-and-family-demographics

U.S. Department of Defense. (2019). *Defense Health Program. Fiscal year (FY) budget estimates 2004-2020*. Washington, DC: Office of the Under Secretary of Defense (Comptroller). Retrieved from https://comptroller.defense.gov/Budget-Materials/FY2020BudgetJustification/#defhealthprog

U.S. Department of Defense, Office of the Secretary of Defense. (2014). *Military health system review. Military health system review—Final report*. Washington, DC: Department of Defense, Secretary of Defense.

U.S. General Accounting Office. (1993). *Performance budgeting: State experiences and implications for the federal government, 12*. Retrieved from http://www.gao.gov/products/AFMD-93-41%0Ahttp://files/1428/AFMD-93-41.html

U.S. General Accounting Office. (1997, March). *Performance budgeting: Past initiatives offer insights for GPRA implementation.* (Publication No. GAO/AIMD/GGD-99-67). Retrieved from https://www.gao.gov/assets/230/223810.pdf

U.S. General Accounting Office. (1999, April). *Performance budgeting: Initial experiences under the results act in linking plans with budgets* (Publication No. GAO/AIMD-97-46). Retrived from https://www.gao.gov/assets/230/227280.pdf

U.S. Government Accountability Office. (2006, December). *Defense health care: Access to care for beneficiaries who have not enrolled in TRICARE's managed care option* (Report No. GAO-07-48). Retrieved from https://www.gao.gov/assets/260/255029.pdf

U.S. Government Accountability Office. (2011, June). *Defense health care: Access to civilian providers under TRICARE standard and extra* (Publication No. GAO-11-500). Retrieved from https://www.gao.gov/assets/320/319092.pdf

U.S. Government Accountability Office. (2013, April). *TRICARE multiyear surveys indicate problems with access to care for nonenrolled beneficiaries.* (Publication No. GAO-13-364). Retrieved from https://www.gao.gov/assets/660/653487.pdf

U.S. Government Accountability Office. (2017, June). *Department of Defense: Actions needed to address five key mission challenges* (Publication No. GAO-17-369). Retrieved from https://www.gao.gov/assets/690/685227.pdf

U.S. Government Accountability Office. (2018a, November). *Defense health care: Additional assessments needed to better ensure an efficient total workforce additional assessments needed to better ensure an efficient total workforce* (Publication No. GAO-19-102). Retrieved from https://www.gao.gov/assets/700/695694.pdf

U.S. Government Accountability Office. (2018b, September). *Managing for results: Government-wide actions needed to improve agencies' use of performance information in decision making*. (Publication No. GAO-18-609SP). Retrieved from https://www.gao.gov/assets/700/694269.pdf

U.S. Government Accountability Office. (2018c, October). *DOD should demonstrate how its plan to transfer the administration of military treatment facilities will improve efficiency*. (Publication No. GAO-19-53). Retrieved from https://www.gao.gov/assets/700/695157.pdf

Van Herck, P., De Smedt, D., Annemans, L., Remmen, R., & Rosenthal, M. B. (2010). Systematic review : Effects, design choices, and context of pay-for-performance in health care. *BMC Health Services Research*, 10(247), 1–14.

VanLare, J. M., & Conway, P. H. (2012). Value-based purchasing—National programs to move from volume to value. *New England Journal of Medicine*, *367*(4), 292–295. https://doi.org/10.1056/NEJMp1207775

Vergun, D. (2013, June). *Soldiers to see fewer PCS moves*. Retrieved from https://www.army.mil/article/104647/soldiers_to_see_fewer_pcs_moves

Vie, L., Griffith, K., Scheier, L., Lester, P., & Seligman, M. (2013). *The person-event data environment: Leveraging big data for studies of psychological strengths in soldiers*. Retrieved from https://www.frontiersin.org/article/10.3389/fpsyg.2013.00934

Vie, L., Scheier, L., Lester, P., Ho, T., Labarthe, D., & Seligman, M. P. (2015). The U.S. Army person-event data environment: A military–civilian big data enterprise. *Big Data*, *3*(2), 67–79. https://doi.org/10.1089/big.2014.0055

Vokes, R. A., Bearman, G., & Bazzoli, G. J. (2018). Hospital-acquired infections under pay-for-performance systems: An administrative perspective on management and change. *Current Infectious Disease Reports*, *20*(35). https://doi.org/10.1007/s11908-018-0638-5

Waldman, J. D., Kelly, F., Arora, S., & Smith, H. L. (2010). The shocking cost of turnover in health care. *Health Care Management Review*, *35*(September), 206–211.

Waters, T. M., Daniels, M. J., Bazzoli, G. J., Perencevich, E., Dunton, N., Staggs, V. S., . . . Shorr, R. I. (2017). Effect of Medicare's nonpayment for hospital-acquired conditions: Lessons for future policy. *JAMA Internal Medicine*, *175*(3), 347–354. https://doi.org/10.1001/jamainternmed.2014.5486.Effect

Waymire, T., & West, T. (2010, December). Performance-based budgeting worth it over long term. *Federal Times*. Retrieved from http://www.federaltimes.com/article/20101205/ADOP06/12050305/

Weech-Maldonado, R., Qaseem, A., & Mkanta, W. (2009). Operating environment and USA nursing homes' participation in the subacute care market: A longitudinal analysis. *Health Services Management Research*, *22*(1), 1–7. https://doi.org/10.1258/hsmr.2008.008002

Weil, A. R. (2019). Military health systems. *Health Affairs*, *38*(8), 1243–1243. https://doi.org/10.1377/hlthaff.2019.00879

Werner, R., Kolstad, J., Stuart, E., & Polsky, D. (2011). The effect of pay-for-performance in hospitals: Lessons for quality improvement. *Health Affairs*, *30*(4), 690–698. https://doi.org/10.1377/hlthaff.2010.1277

Werner, R. M., & Dudley, A. (2012). Medicare's new hospital value-based purchasing program is likely to have only a small impact on hospital payments. *Health Affairs*, *31*(9), 1932–1940. https://doi.org/10.1377/hlthaff.2011.0990

Werner, R. M., Konetzka, R. T., & Polsky, D. (2013). The effect of pay-for-performance in nursing homes: Evidence from state medicaid programs. *Health Services Research*, *48*(4), 1393–1414. https://doi.org/10.1111/1475-6773.12035

West, T., & Cronk, M. (2011). Financial officers as leaders: Lessons from Army medicine. *Strategic Finance*, *92*(10), 31–37. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=59789978&site=ehost-live

West, T., Cronk, M., Goodman, R., & Waymire, T. R. (2010). Increasing accountability through performance-based budgeting. *Journal of Governement Financial Management*, *Spring*, 51–55.

Weyer, S. M., Bobiak, S., & Stange, K. C. (2008). Possible unintended consequences of a focus on performance: Insights over time from the Research Association of Practices Network. *Quality Management in Health Care*, *17*(1), 47–52. https://doi.org/10.1097/01.QMH.0000308637.04850.ac

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, *48*, 817–830.

Williams, C., & Melhuish, W. (1999). Is ABCM destined for success or failure in the federal governement? *Public Budgeting and Finance*, *19*(2), 22–36.

Winslow, B. R. (1994, September 6). HMO juggernaut: U.S. healthcare cuts costs, grows rapidly and irks some doctors—A few patients are slighted, squeezed specialists say: But firm also gets praise—How Katie avoids hospital. *Wall Street Journal*, p. A1. Retrieved from http://proxy.library.vcu.edu/login?url=https://search.proquest.com/docview/398560281?accountid=14780

Wooldridge, J. M. (2013). *Introductory econometrics: A modern approach* (5th ed.). Mason, OH: South-Western College.

Yeager, V. A., Menachemi, N., Savage, G. T., Ginter, P. M., Sen, B. P., & Beitsch, L. M. (2014). Using resource dependency theory to measure the environment in health care organizational studies: A systematic review of the literature. *Health Care Management Review*, *39*(1), 50–65. https://doi.org/10.1097/HMR.0b013e3182826624

Yeager, V. A., Savage, G. T., Ginter, P. M., & Beitsch, L. M. (2014). Using resource dependency theory to measure the environment in health care organizational studies : A systematic review of the literature. Health Care Management Review, 39(1), 50–65. https://doi.org/10.1097/HMR.0b013e3182826624

Yip, W., Powell-Jackson, T., Chen, W., Hu, M., Fe, E., Hu, M., . . . Hsiao, W. C. (2017). *Capitation combined with pay-for-performance improves antibiotic prescribing practices in rural China*. https://doi.org/10.1377/HLTHAFF.2013.0702

Zemrani, A. (2019). Budgeting in the United States: From theory to practice using higher education. In M. S. DeVries & J. Nemec (Eds.), *Performance-based budgeting in the public sector* (pp. 29–60). https://doi.org/10.1007/978-3-030-02077-4_2

Zhan, C., & Miller, M. (2003). Excess Length of Stay, Charges, and Mortality Attributable to Medical Injuries During Hospitalization. *Journal of the American Medical Association*, *290*(14).

Zhang, N., Li, Y., & Temkin-Greener, H. (2013). Prevalence of obesity in New York nursing homes: Associations with facility characteristics. *Gerontologist*, *53*(4), 567–581. https://doi.org/10.1093/geront/gnt011

Zimmerman, J. (2003). *Accounting for decision making and control* (4th ed.). Boston, MA: McGraw-Hill Irwin.

Zinn, J. S. (1994). Market competition and the quality of nursing home care. *Journal of Health Politics, Policy and Law*, *19*(3), 555–582. https://doi.org/10.1215/03616878-19-3-555

Zinn, J. S., Proenca, J., & Rosko, M. D. (1997). Organizational and environmental factors in hospital alliance membership and contract management: A resource-dependence perspective. *Hospital & Health Services Administration*, *42*(1), 67–86.

Zinn, J. S., Weech, R. J., & Brannon, D. (1998). Resource dependence and institutional elements in nursing home TQM adoption. *Health Services Research*, *33*(2 Pt 1), 261–273. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1070264&tool=pmcentrez&rendertype=abstract

Vita

Kimberly Decker was born on October 7, 1981 in Hampton. Virginia. She graduated from the University of Virginia in 2004 with a bachelor's degree in psychology. While serving in the U.S. Army as a Medical Service Corps officer, she received a Master of Health Administration degree from Baylor University in 2012. During her time in the Army, she has worked as a healthcare administrator at Irwin Army Community Hospital, Martin Army Community Hospital, Munson Army Health Clinic, Winn Army Community Hospital, and Tuttle Army Health Clinic. She has been a Fellow of the American College of Healthcare Executives since 2014 and holds a Black Belt certification in Lean Six Sigma. As of the summer of 2020, she is an instructor at the Army-Baylor University Graduate Program in Health and Business Administration.