ENGELBERG CENTER for
Health Care Reform
at BROOKINGS

Evaluating Community-Based Reforms in Care for Chronic Conditions

A Multi-Payer Template for Information Technology Initiatives

Table of Contents

Executive Summary

The pressure for rapid and effective change in health care delivery in the United States has continued to intensify as health care costs continue to rise. Communities have responded with a broad set of initiatives designed to improve the quality and reduce the cost of health care. Many of these initiatives have focused on improving care for those with chronic conditions—which account for the vast majority of health care spending—and health information technology (HIT) has been a fundamental enabler of many of them.

Community-level interventions have the advantage of facilitating system-wide delivery and payment reform because they shift the focus of reform efforts away from particular institutions to the broader community where patients receive their care and manage their conditions, thus facilitating greater care coordination across settings. Community-based initiatives, however, have created a new set of methodological challenges around how best to determine whether and to what extent these initiatives have had an impact on cost and quality and, where they have, determining what worked and what could have worked better.

The challenge of evaluation derives from the nature of community-based innovation. Initiative design and primary objectives may vary in different communities, given different priorities and the types of available resources; typically lack randomization and control in an environment where patients may elect to participate in innovative programs non-randomly based upon characteristics that affect measured health outcomes; proceed in a dynamic environment, where many things may be changing simultaneously; and they may have differential, and sometimes completely opposite effects, on subsets of the population, in a setting where there is very limited opportunity to evaluate the impact on some of those subsets. The risk in that setting is that apparent improvements in an outcome, such as cost, in one population (e.g., Medicare enrollees) may be the consequence of cost-shifting to an unobserved population (e.g., commercial enrollees) rather than the result of genuine improvements in the efficiency with which care is delivered.

To begin to address these challenges, the Engelberg Center for Health Care Reform at Brookings has worked closely with three communities, the Primary Care Information Project (PCIP) in New York City, Vermont's Blueprint for Health (Blueprint) and the Wisconsin Health Information Organization (WHIO) , and analysts from the Dartmouth Institute, New York University, Onpoint Health Data and OptumInsight (formerly Ingenix) to develop a multi-payer template for evaluating their ongoing and future community-based health care reform initiatives. The objective was to better equip these communities to conduct rigorous evaluation of the impact of their interventions on the cost, quality, and utilization of health care and to provide a template that would enable other communities implementing accountability-based multi-payer payment and delivery system reform initiatives to understand how they might do so.

Methodology

The basic approach to developing a multi-payer evaluation template at the community level involved a series of steps intended to permit sites to establish a reliable baseline for purposes of planning for, and evaluating the future impact of, reform initiatives. The sites included PCIP, a community effort in New York City to improve access to timely preventive services in the ambulatory care setting through electronic health record electronic health record adoption and data exchange, with a particular focus on the prevention of chronic illness among underserved populations; Vermont's Blueprint, a statewide initiative using patient centered medical homes to reduce health care costs through improved care

coordination, with a focus on patients with chronic conditions; and WHIO, a statewide initiative that uses an all-payer database to improve the quality, affordability, safety, and efficiency of health care. The steps to developing a multi-payer evaluation template in these communities involved the following:

- Identifying multi-payer sources for data to support evaluation;
- Identifying and prioritizing a set of measures that addressed cost, quality, and utilization and that were relevant to each community's priorities for health care improvement;
- Standardizing measure specifications for the selected measures across data sources;
- Choosing a level of analysis (e.g., intervention level for PCIP and Blueprint and county-level for WHIO) and defining pools of populations for comparison (e.g., comparison and intervention groups for PCIP and Blueprint and counties for WHIO);
- Selecting and implementing methods for attributing patients to one or the other of those populations, as appropriate;
- Selecting and implementing methods to describe and account for differences in patient risk, across those populations; and
- Estimating those metrics for the appropriate level of analysis for each site.

Summary of Findings

In general, these steps were implemented so that baseline values for critical metrics could be established in all three communities. In New York City (PCIP) and Vermont (Blueprint), the ability to develop a multi-payer baseline was an essential step toward preparing to evaluate the impact of health reform initiatives that are underway. As a result, critical questions related to the extent to which "intervention" populations were similar to (or different from) "comparison" (non-intervention) populations at baseline. Through the steps above, baseline differences in patient characteristics, utilization, cost, and quality of care within insured groups (e.g., between Medicare enrollees in the "intervention" population and those in the "comparison" group) were evaluated in PCIP and Blueprint. Recognizing, and then accounting for, these differences should permit the impact of reform initiatives to be isolated in analyses going forward.

In Wisconsin, that baseline revealed variation across counties and between payers within counties. Notable differences in the apparent "value" of care across counties in Wisconsin appeared when total cost was examined against two quality metrics (ambulatory care sensitive hospital admissions and receipt of three recommended diabetes tests). Eighteen counties (clustered mainly in the mid-eastern region of the state and a small group along the middle of the western border) appeared to provide health care that was better quality and lower cost than the statewide payer average across two or more payers, while five counties appeared to provide lower quality and higher cost health care than the statewide payer average across two or more payers. While further tests may need to be run to determine the statistical significance of this variation, these differences may be important to policy makers in Wisconsin, as they consider where to focus resources to improve health and health care.

Systematic efforts to adjust for risk using a variety of tiered risk-adjustment methods had relatively little impact on population level results in each of the three communities. Communities may want to invest in risk-adjustment, but modest investments—for example, in methods and data collection to enable for basic adjustments for age, race, and gender—may be sufficient in communities where evaluation resources are limited.

Key Challenges and Lessons Learned in Building a Community-Level Multi-Payer Evaluation Template
In the process of developing multi-payer community-based evaluation baselines for each of the three communities, several challenges and key lessons emerged:

- Developing a common template can be challenging in light of community priorities and constraints: The three communities differed in overall health objectives to the nature of their reform initiatives and maturity of their interventions. Because of PCIP's focus on underserved populations, this study analyzed the Medicare and Medicaid populations in the New York City area. Blueprint is a statewide initiative that rolled out at different sites at different times, and due to concerns around staggered start-up dates and sample size, each site in the commercial population (Burlington and St. Johnsbury) was evaluated independently, while the Medicare data evaluated three sites together (Barre, Burlington, and St. Johnsbury) using time-adjusted results. WHIO maintains an All Payer Claims Database (APCD) that uses a multi-payer data mart to aggregate data and produce provider performance measure reports with the goal of improving quality, affordability, safety, and efficiency of health care. APCD permitted statewide analysis at the county level across all payer populations—Medicare, Medicare Advantage, Medicaid, and commercial. Even given a workable common template, these differences required that approaches to developing baselines be tailored to each of the three communities.
- Because value cannot be assessed by quality, utilization or cost alone, all three dimensions should be looked at concurrently: Significant but inconsistent variations in cost, utilization, and quality across populations were observed, which suggests it will be difficult to predict with confidence ahead of time how an intervention designed to improve quality will affect cost (or how an intervention designed to improve efficiency might affect quality). This calls for multi-dimensional assessment and may be especially important where outcomes may vary across covered populations (e.g., where costs may be shifted from one covered group to another).
- Measuring total cost of care in standardized ways across payers can be challenging but is critical for health reform: Given the importance of cost as an outcome of health reform, a viable measure to assess total cost of care at each site needed to be developed. This was difficult because not all sites had access to data on allowed charges and because standardized costs data may not accurately reflect local prices. Variability of data, inconsistent use of provider identifiers, differences in benefit design, changes in enrollment, and lack of good price data created challenges in crafting a standardized total cost of care measure.
- Distributed approaches to evaluation are viable and can be efficient: Differences in the structure and quality of data across payers can make analysis across payers challenging. Anticipating these challenges, this project relied on a distributed approach to data analysis, which minimized the transfer of raw patient-level data and enabled the data to remain closer to the source. Though this distributed approach required extensive coordination, the approach facilitated timely analysis of the data because it made those individuals who knew the data best responsible for cleaning and standardizing the data and running the measure specifications against that data, with summary data being exchanged with Brookings.
- Access to health outcome data would have made these evaluation templates stronger, however claims-based measurement can be highly effective and meaningful: Claims measures are very useful—particularly for measuring utilization and total cost of care—but are limited in their clinical utility. Over time, and as HIT adoption increases, community initiatives should move toward clinically-enriched and advanced measures so that health outcomes can also be assessed. Nevertheless, this project has demonstrated that low HIT adoption rates need not be a barrier to reform implementation, as meaningful and effective performance measurement is possible at every stage of HIT adoption.

- New evaluation approaches that account for the lack of randomization and control may be needed to permit the impact of policy changes to be evaluated more finely: Unlike scientific experiments, which take place in carefully controlled environments and test for the marginal effects of isolated interventions, many changes took place simultaneously in all of the communities in which this work was done. The baselines determined through the applied methods will be a critical resource to those who hope to determine, going forward, whether cost and health outcomes have improved. Challenges will exist trying to determine, in each community, what policy reforms might be most important to accounting for that change.

Preparing for an Era of Greater Accountability

The changes that are taking place in communities like those participating in this study are becoming both more urgent and more widespread. There will continue to be movement away from fee-for-service payments (based on the volume and intensity of services regardless of their quality) to payment reforms that demand more accountability for cost and quality. Communities, like the ones here, that begin to plan for the transition to accountability-based payments are likely to find the transition both more orderly and more effective. A critical part of that planning is around evaluation: how will communities know that change has led to improvement?

This project was intended to facilitate reform efforts through the development of sound evaluation baselines against which future impacts can be assessed and can, in turn, be used to identify effective community-based HIT-enabled interventions and encourage their wider dissemination. While developing such an infrastructure is a resource-intensive endeavor, communities should be able to realize meaningful care improvements that result in cost savings and create resources that can then be re-invested into the community to help finance such investments. Incremental steps like these are critical for initiating a virtuous cycle of quality improvement activities that can help finance further investments in the infrastructure needed to sustain them and are necessary steps along the way toward payment systems that ultimately support providers and communities when they do the right thing for patients.

Overview of the Report

This summary provides a very brief summary of work that is detailed much more completely in the accompanying report, *Evaluating C ommunity-Based Reforms in Care for Chronic Conditions: A Multi-Payer Template for Information Technology Initiatives*:

- An **introduction** that provides important background, including selection criteria for the sites and detailed descriptions of the sites and of the health reform initiatives underway at each;
- A **methodology** section that provides much more detail about the steps taken to develop the template, and about the specific techniques used to address issues such as risk-adjustment and patient attribution;
- A **study findings section** that presents in great detail the baseline characteristics of intervention and comparison populations in PCIP and Blueprint, a county-level assessment of results in WHIO, and an assessment of the impact of risk adjustment on results for each site, a cross-site Medicare analysis, and a summary of key findings;
- A **discussion** section, in which important lessons learned in building an evaluation template are extrapolated; and
- A **conclusion** with recommendations for future evaluations of health care reform effort.

Chapter 1
Introduction

Chapter 1: Introduction

Despite high and rising health care costs, Americans often do not get the care they need. Recent studies have identified many areas of apparent waste.[1] Simultaneously, many important services are underused, and adherence to proven-effective therapies for many chronic illnesses remains low.[2,3,4] One of the most significant examples of the gap between the care that Americans receive and the quality of care that is possible lies in the potential for health information technology (HIT) to improve results for patients with chronic disease. People with multiple chronic conditions are among the highest users of medical services. One study found that the cost of care roughly doubles with the addition of each chronic condition.[5] Another study found that the one quarter of Medicare beneficiaries with five or more chronic conditions account for 70 percent of total program costs.[6] Thus, chronic disease has emerged as the most widespread and costly health problem in the United States and an opportunity for significant health care savings.

In response to this challenge, a number of multi-payer community-and regionally-based initiatives are being developed to combine payment and delivery reform, through the effective use of HIT, to improve care coordination and management of chronic disease. Addressing the challenge of effective chronic disease care must be a core focus of any meaningful health care reform initiative. Such health care reform approaches may be the only way to reduce costs without compromising quality or losing public support for reform. More effective treatment of chronic disease will require care coordination across multiple care settings, including inpatient, outpatient, and home care. Coordination across such a complex environment will require better information systems that allow for the effective and timely exchange of data. Improved information technology can also help transition the locus of care from the expensive and often dangerous environment of the hospital to the community where social support resources can be more effectively deployed.

With funding from the Markle Foundation and assistance from the Dartmouth Institute (Dartmouth), the Engelberg Center for Health Care Reform at Brookings (Brookings) has worked closely with three community-based initiatives that either have implemented—or are preparing to embark on—a range of HIT-enabled delivery and payment reforms: the Primary Care Information Project (PCIP) in New York City, the Wisconsin Health Information Organization (WHIO), and Vermont's Blueprint for Health (Blueprint). The objectives of this project were to (1) assist the communities in developing a common core set of multi-payer quality, utilization, and cost metrics; (2) formulate a baseline against which each community can assess the impact of their initiatives moving forward and identify areas where future reform efforts can be targeted; and (3) facilitate the comparative analysis of interventions across the three communities in order to accelerate the sharing of best practices. In doing so, this template could further enable other communities implementing accountability-based multi-payer payment and delivery system reform initiatives to understand how they might also conduct rigorous evaluation of the impact of their interventions on the cost, quality, and utilization of health care.

---

[1] Thomson Reuters. (October 2009). Where Can $700 Billion in Waste Be Cut Annually from the US Healthcare System?.
[2] McGlynn EA, et al. (2003). The Quality of Health Care Delivered to Adults in the United States. *New England Journal of Medicine* 348: 2635–45.
[3] Agency for Healthcare Research and Quality. *National Health Care Quality Report, 2008*. (2008). Rockville, MD.
[4] Elixhauser A and Owens P. *Adverse Drug Events in U.S. Hospitals, 2004. Healthcare Cost and Utilization Project Statistical Brief 29*. (2007). Agency for Healthcare Quality and Research.
[5] Anderson A, Horvath J (2004). The Growing Burden of Chronic Disease in America. *Public Health Reports* 119 (3): 263-270
[6] Anderson G. (2005) Medicare and Chronic Conditions. *New England Journal of Medicine,* 353(3): 305-309

This report begins with a brief overview of the motivation behind this project. A brief description of the three communities that participated and a description of the methods used to derive evaluation baselines for each community follow. The report discusses key findings and concludes with a discussion of what the implications might be for evaluating delivery and payment reforms at the community level. Challenges encountered and lessons learned across the project sites are also described to inform future community-based health reform initiatives.

Project Overview: A Coordinated but Distributed Approach to Evaluation

Despite widespread agreement on the need for a comprehensive, community-wide, multi-payer approach to delivery and payment reform, little evidence exists on how to effectively evaluate such initiatives. Because effective reform involves interventions across multiple dimensions, such as care coordination, payment reform, and delivery redesign, isolating the effect of any one component is difficult. Local factors such as patient demographics and market concentration can also affect the outcome of an intervention. Thus, a common core set of consistent measures of quality, cost, and utilization that can be used consistently across multiple payers is a key component to effective evaluation. Specifying these multi-payer core metrics at the beginning promotes clarity about what the objectives are, which facilitates alignment around them across payers, promoting signal strength.

The goal of this project was to assist in the development of better evidence on how payment reforms and HIT can be used to improve quality and reduce the cost of care for chronically ill individuals. As part of this project, Brookings has worked closely with the PCIP, WHIO, and Blueprint to develop a pre-intervention multi-payer baseline against which each community can, moving forward, measure the effects of their interventions.

Brookings worked with the Dartmouth Institute, New York University, Onpoint Health Data and OptumInsight (formerly Ingenix) and the three communities to develop a consistent but distributed approach to evaluation. While the process of identifying and selecting a common core set of consistent measures and agreeing upon relatively consistent provider matching, patient attribution, and risk adjustment techniques were centrally coordinated through Brookings, the raw data remained with the communities (and with Dartmouth in the case of Medicare data). The data were managed and cleaned by the data holders and the data methods were applied locally, with only summary data exchanged with Brookings for use in preparing this report.

Project Site Selection Criteria

The following selection criteria were employed to identify the most appropriate sites for this project:

- Multi-payer involvement: The number of payers participating and the scope to build “signal strength” to elicit positive change in care design and delivery;
- Potential for payment reforms: The potential and feasibility of using coordinated payment reforms to drive improvements in chronic disease coordination and care to expand the reach and impact of these efforts;
- Advanced IT infrastructure: The extent of existing HIT capabilities and the degree to which HIT is integrated with delivery/payment reform efforts;
- Community-based: The level of community involvement and multi-stakeholder participation, for example, a collaborative, community-based approach that is not centred on any one particular organization or locus of care.

Site Descriptions

Primary Care Information Project—New York City, New York

Overview: PCIP is a city-wide initiative and a Bureau of the New York City Department of Health and Mental Hygiene (DOHMH). PCIP's key health objective is to improve access to timely preventive services in the ambulatory care setting, with a particular focus on primary and secondary prevention of chronic illnesses among underserved communities. The program involves DOHMH staff and a variety of the primary care settings in New York's medically underserved communities. The goal of PCIP is to improve the health of New York City's medically underserved population through the use of HIT and data exchange.

In 2005, Mayor Bloomberg created PCIP as a priority health initiative for New York City. Over the years, PCIP has grown into a 90-person bureau with a cumulative budget of $60 million in city, state, federal, and private funds. The program seeks to spur electronic health record (her) adoption by primary care providers serving the city's disadvantaged communities. Program staff help local primary care doctors adopt EHRs and connect to health information exchange (HIE) networks. EHRs adopted by PCIP automatically generate and send quality measures to a citywide data repository. These measures are derived from the city's 10 priority health areas and are focused on prevention.[7]

In January 2009, PCIP launched NYC Health eHearts Rewards, a pilot quality reporting and incentives program funded by the Robin Hood Foundation, because they recognized that EHR adoption alone would not improve quality of care. EHR-enabled providers participating in the pilot program send reports on a core set of cardiovascular measures (related to aspirin use, blood pressure control, cholesterol control, and smoking cessation interventions) to PCIP's data repository. These measures were selected for potentially having the most impact for reducing avoidable death caused by heart attacks and stroke.[8] Half of the participating practices were randomized to receive reimbursement for meeting recommended clinical goals on these measures. These measures are a subset of the Adult Primary Care Take Care New York measures already in place in PCIP when this project began, which focus on the delivery of preventive services.

IT Infrastructure: A majority of practices that have adopted an EHR use the eClinicalWorks (eCW) software co-developed with PCIP to incorporate the Take Care New York (TCNY) measures, with features designed to improve population health. These features include a clinical decision support system, which issues preventive care reminders and helps practices adhere to clinical guidelines; a registry function, which allows providers to rapidly identify cases of interest and track their performance in managing patients with specific conditions; and an automated system for creating quality measure reports, which offer a practice-wide snapshot of patient performance on the TCNY measures.

Data from practices using the TCNY build of eCW transmit data to PCIP's data warehouse, the Healthcare Quality Information Network (HQIN). EHR users input data into their patient records, and software programs integrated at the practice calculate measures in a standardized format that is then transmitted through eCW to HQIN. The original data remains with the practice and only specific summary sets are extracted through automated queries embedded in the EHR software. The data collected by PCIP is an

[7] Retrieved December 13, 2011, from http://www.nyc.gov/html/doh/downloads/pdf/tcny/tcny-2012.pdf

[8] Farley, et. Al. "Deaths Preventable in the U.S. by Improvements in the Use of Clinical Preventive Services." Elsevier Editorial System for American Journal of Preventive Medicine. Retrieved December 13, 2011, from http://www.preparedpatientforum.org/research/support_050410.pdf.

example of distributed or federated data architecture because only limited information is extracted and transmitted (See Exhibit 1 for description of centralized or distributed data architectures).

New York DOHMH monitors these summary measures for population health surveillance purposes and utilizes the subset to assess the performance of providers participating in the NYC Health eHearts program. PCIP also generates a practice-specific score card that summarizes quality improvement performance. These score cards are shared with the providers and technical assistance is provided by quality improvement field staff trained by PCIP. Additional feedback loops to providers include in-person visits from PCIP quality improvement specialists to practices, and showing providers how to access similar information to the score cards within the EHR with the quality report tool feature or clinical decisions support system.

PCIP worked with the Health Care Incentives Improvement Institute, an organization that develops standardized score cards. In collaboration with the Institute's Bridges to Excellence (BTE) program, PCIP developed the Adult Primary Care Recognition Program scores for practitioners. This score card uses data aggregated from EHRs by PCIP, and with permission from the providers, transmits the necessary summaries to BTE for scoring. Providers that meet the scoring requirements are publicly recognized. BTE then shares the provider's recognition status with insurers or employers to determine eligibility of financial rewards or quality distinctions that they may offer.

Comparing participating practices with non-participating practices would be necessary in order to evaluate the effectiveness of PCIP. However, non-participating practices may not have the same EHR capabilities to generate these performance measures. The claims-based quality, cost, and utilization measures developed as part of this project were particularly important to the PCIP program. To the extent PCIP and Health eHearts aimed to improve health outcomes through secondary prevention, the ability to track reductions in emergency department (ED) and hospital use for ambulatory care sensitive conditions was particularly important. Finally, being able to generate these measures using Medicare and Medicaid claims data was also important because PCIP targets underserved and vulnerable patients.

Vermont Blueprint for Health—State of Vermont

Overview: Blueprint is a statewide, public–private initiative designed to reduce the health and economic impacts of common chronic conditions. The initiative takes a systems-based approach to health care transformation. The Blueprint Integrated Pilot model will be implemented and evaluated in three distinct communities across Vermont as part of the effort. The first two pilot communities (Burlington and St. Johnsbury health service areas (HSAs)) began operations in 2008; a third was launched in 2009 in Barre.

The key health objectives of Blueprint are to reduce health care costs through effective primary and secondary prevention of common chronic conditions and improved care coordination. A key component of the pilots is a multidisciplinary Community Care Team (CCT) that provides support and expertise to participating medical practices through direct services, care coordination, population management, and quality improvement activities. HIT and statewide prevention campaigns support these community-based efforts. The pilots test a public-private approach to state-initiated multi-payer reform in which primary care practices employing Patient Centered Medical Homes are paid a per-patient-per-month bonus according to their quality of care as measured by the Physician Practice Connections®–Patient-Centered Medical Home™ program, operated by the National Committee for Quality Assurance (NCQA). Performance measures include NCQA Patient-Centered Medical Home scores, guideline based care for

chronic conditions and health maintenance, improved control in chronic health conditions, and outcomes data such as morbidity and mortality.

To financially support Vermont's multi-payer efforts in improving care quality, the project specifically aims to reduce healthcare expenditures associated with avoidable hospitalizations and emergency department visits. The multi-payer cost and utilization measures developed as part of this project, particularly those that assess hospital and ED use for ambulatory-care sensitive conditions, were especially important to the Blueprint project.

Each of the pilot's CCTs includes clinical staff selected to address community needs and a public health prevention specialist based in the local Department of Health District Office. The prevention specialist works with other CCT personnel to guide quantitative and qualitative community evaluation of the risk factors and conditions contributing to the prevalence and morbidity from chronic disease. In addition, CCTs build consensus among local health care stakeholders on priority areas for public health interventions. Finally, CCTs are responsible for implementing interventions in collaboration with community stakeholders and evaluate their effect on the prevalence and impact of chronic disease.

IT Infrastructure: A major priority of Vermont Blueprint is the development of HIT infrastructure that will facilitate improvements in cost and quality of care for people with chronic diseases as well as statewide reporting to track the health of the population. At no cost to health care providers, the state is offering a web-based clinical system that informs treatment decisions, facilitates electronic prescribing, and enables practice-level reporting of data on patient populations and outcomes to the state.

Vermont Blueprint works closely with the Vermont Information Technology Leaders (VITL)—the state—sponsored HIE—to develop infrastructure to support the program interventions. The core of this infrastructure is the Blueprint's centralized registry and web-based clinical tracking system, DocSite-Covisint. The registry is used to produce visit planners that guide individual patient care, and to produce reports that support population management, quality improvement, program evaluation, and comparative benchmarking. The system also establishes a repository of common data elements, enabling communities and the state to track population health and evaluate the effectiveness of improvement interventions. To monitor system costs and utilization, this clinical registry is complemented by a multi-payer commercial claims database managed by Onpoint.

Unlike PCIP distributed data architecture, Blueprint decided to pursue a centralized model for data collection. Patient-level data are sent from the point of care to the DocSite web-based registry and clinical tracking system, either by entering data manually into the web-based portal or via interfaces and direct feeds. Similarly, health plans submit their claims data to Onpoint for data cleaning and aggregation. Plans are currently underway to further centralize and streamline this data by developing an informatics platform housed at the University of Vermont, which will combine the claims data and the clinical data into one large dataset.

Wisconsin Health Information Organization—State of Wisconsin

Overview: The Wisconsin Health Information Organization is an All Payer Claims Database (APCD) focused on improving quality, affordability, safety and efficiency of health care. The goal of WHIO is to aggregate data and produce provider performance measure reports to examine variations in efficiency, quality, safety, and cost. To achieve this objective, WHIO has created a multi-payer data mart that includes de-identified beneficiary eligibility data; provider data, including specialty assignment and

primary practice location; and 27 months of service-level claims data from 17 of Wisconsin's major payers (including Medicaid). Patient data is de-identified, though provider-level data is not. The data includes claims records details, spans of enrollment, specialty categorizations, and provider and member IDs that have been standardized across payers. The data also includes value-added fields like bills on inpatient stay. Services related to inpatient stay are tagged. Practice groups are created in the data through the use of a Wisconsin Medical Society crosswalk that attaches physicians to practices.

WHIO has convened a multi-stakeholder effort called the Wisconsin Partnership for Healthcare Payment Reform to develop a vision for alternative payment mechanisms in Wisconsin that recognize and reward high value health care providers. The short-term goals of the commission include bundled payment for total knee replacements and improved management of diabetic patients including reducing hospital readmissions and ambulatory-care sensitive admissions and developing benefit designs and processes for public reporting on quality and costs to encourage patients to choose high-value treatments and providers. The commission hopes to develop long-term global payments to medical homes for managing patients' chronic conditions.

IT Infrastructure: WHIO uses Ingenix's Impact Intelligence software and Episode Treatment Groups. Ingenix's Impact Intelligence tool applies quality measures (e.g., Healthcare Effectivness Data and Information Set (HEDIS) and Physician Consortium for Performance Improvement (PCPI)) and risk scores to the standardized WHIO data. The data mart also features a reporting system that displays comparative physician-level results in table format. WHIO produces both physician- and practice-level reports by specialty. WHIO's reports use claims-based measures that are calculated at the identified provider-group level. WHIO eventually hopes to use provider-group level measures as a basis for payment reform efforts.

Having developed one of the most comprehensive pools of health claims information available anywhere in the United States, WHIO is a paradigm example of centralized data architecture. The Exchange holds a rolling 27 months of claims data that comprises the experiences of more than 3.7 million people and 233.5 million treatment services. A total of 21.5 million episodes of care are now found in the database and its scope will grow as new members join and contribute their data. This large and growing pool of data will support an information marketplace where members can access data and receive reports that analyze health system and physician performance based on hundreds of variables. Hopefully, these reports can be used to identify gaps in care for treatment of chronic conditions, costs per episode of care, population health, preventable hospital readmissions, and variations in generic prescribing.

Exhibit 1: Centralized vs. Distributed Data Architectures

Data architectures for HIT systems can be divided into two broad categories: centralized and distributed (also referred to as federated of decentralized). Although there are many variations within each category, the main factor separating centralized from distributed architectures is the extent to which patient level data is transferred from the data owner to a central data repository or aggregator.

	Centralized Model	Distributed Model
Overview	Copies of original data are collected from multiple sources, usually in different formats, and brought together at one centralized warehouse. Collected data usually includes detailed patient data including personally identifiable information, clinical data, social history, and demographic data. One way to conceptualize a centralized model is that 'data is brought to the question': that is, data collection is usually undertaken to answer specific research questions, leading to the creation of single-purpose datasets.	Local control of data is maintained as data stays with the original source, behind the data owner's institutional firewalls. Instead of sending exact copies of the data as in the centralized model, an image of the data is created in standardized format to ensure identical file structure and data element definitions across data sources. Queries can then be run against the standardized data images to generate summary-level results (numerator, denominator, rate information), which can then be aggregated for various analytical purposes. In this way a distributed model can be thought to 'bring the question to the data' in that local control of the data is maintained, analytics are performed locally, and only summary-level results are transmitted.
Examples	National systems such as the CMS Integrated Data repository; state and regional models such as Blueprint's DocSite or the WHIO all payer claims database.	AHRQ Distributed Research Network; The Distributed Surveillance Taskforce for Real-time Influenza Burden Tracking (DiSTRIBuTE); the Food and Drug Administration's Mini-Sentinel adverse drug event surveillance system; and the TCNY-QRS used by the PCIP in New York.
Pros and Cons	Centralized models can generate data sets to answer narrowly and broadly defined questions. Some may prefer the centralized model because it seems more intuitive since data is aggregated and warehoused centrally. However, there are also drawbacks to data systems with centralized architectures. Since sensitive, patient-level data is transmitted and pooled, centralized models carry higher privacy and security risks. In addition, continual requirements to report single data sets can create reporting burden on providers. Finally, analyzing aggregated data is often difficult and time consuming as the data often exists in many formats and must be cleaned and normalized before analysis can commence, often by third-parties who may not be very familiar with the idiosyncrasies of the data they receive.	Since no identifiable patient data is transmitted outside the data owner, many patient privacy and security concerns are allayed. The flexibility of a distributed system means that one data infrastructure can be used for multiple uses. Lastly, data images are standardized by the data holders most familiar with any idiosyncrasies of the data, ensuring that analysis of the data is timely and complete. The main drawbacks of a decentralized system are complexity and the need for continual data refreshes and coordination across data holders to ensure consistency. Drawbacks to this model include the substantial infrastructure investment required to build a distributed database system and the site-level capacity needed to maintain the data.

Chapter 2
Methodology

Chapter 2: Methodology

This chapter describes the methods used to develop a pre-intervention baseline for Vermont's Blueprint and New York City's PCIP and the methods used to provide Wisconsin's WHIO. A statewide snapshot of county-level variations in quality and cost of care to aid WHIO in identifying promising areas to target their future payment and delivery reform interventions is also included in this chapter.

Each baseline consisted of a set of performance measures that could be applied to both Medicare and non-Medicare data. To the extent possible, the measures were standardized across sites, however, due to funding and resource constraints, some sites prioritized certain measures over others. For Vermont and New York, developing the baseline also consisted of allocating physicians to "intervention groups" and "comparison groups" by identifying the physicians who are participating in either the Blueprint or PCIP program and attributing patients to them.

This chapter describes the following steps taken to develop these baseline evaluations for each site:

- Identifying multi-payer sources for data to support evaluation;
- Identifying and prioritizing a set of measures that addressed cost, quality, and utilization *and* that were relevant to each community's priorities for health care improvement;
- Standardizing measure specifications for the selected measures across data sources;
- Choosing a level of analysis (e.g., intervention level for PCIP and Blueprint and county-level for WHIO) and defining pools of populations for comparison (e.g., comparison and intervention groups for PCIP and VT and counties for WHIO);
- Selecting and implementing methods for attributing patients to one or the other of those populations, as appropriate;
- Selecting and implementing methods to describe and account for differences in patient risk, across those populations; and
- Estimating those metrics for the appropriate level of analysis for each site.

Each step will be discussed in more detail and a summary is provided in Exhibit 2.

Exhibit 2. Cross-Site Methodological Summary

	PCIP	Blueprint	WHIO
Stage of Site intervention, Overall Objective, and Overview	**Overall Objective:** Improve population health through health information technology and data exchange. **Overview:** Program staff helps practices that serve largely Medicaid and uninsured patients adopt EHRs and connect to health information exchange networks. The EHRs adopted by PCIP practices automatically generate and send quality measures—many of which have a preventive-care focus—to a city-wide data repository. Since January 2009, PCIP has been conducting a pilot performance incentives program.	**Overall Objective:** Implement a primary care medical home model that improves access to well-coordinated preventive health services, centered on the needs of patients and families, in order to reduce avoidable complications from chronic conditions through improved disease control and prevention, and coordinated access to the range of support services that target common contributors to poorly controlled disease and thereby reduce the rate at which health care costs are growing and demonstrate financial sustainability through multi-insurer payment reform and a public-private partnership that results. **Overview:** Intervention practices receive assistance in managing their chronically ill patients from a local, multi-disciplinary Chronic Care Team. Practices also receive access to DocSite, an online clinical tracking tool. Intervention practices are paid a per-patient-per-month (PMPM) bonus based on their scores on the NCQA PPC-PCMH tool.	**Overall Objective:** Wisconsin is in the delivery and payment reform planning stages. To help identify promising areas for improvement, WHIO is assessing county-level variations in cost, quality, and utilization. **Overview:** In April 2010, WHIO convened a multi-stakeholder payment reform summit to begin to develop consensus on the structure of payment reforms to be implemented in Wisconsin. The recommendations will form the basis for work plans that will be created by several volunteer work groups.
Intervention Launch	2008	2008	Wisconsin is still in the process of planning their intervention
Medicare Fee-for-Service (FFS) data source	The Dartmouth Institute	The Dartmouth Institute	The Dartmouth Institute
Identification of non-FFS Medicare data source	New York State Medicaid data set, analyzed by Tod Mijanovich at NYU	Multi-payer data set, managed by Onpoint, containing data from commercial health plans.	WHIO data set, managed by Ingenix, including data from commercial health plans, Medicaid, and Medicare Advantage.
Years of data to be used (baseline)	**Medicaid**: 2007 **Medicare**: 2005-2007 (20 percent sample)	**Commercial**: July 1, 2007 to June 30, 2008 for St. Johnsbury; Oct. 1, 2007 to Sept. 30, 2008 for Burlington **Medicare**: 2005-2007 (20 percent sample) for St. Johnsbury, Burlington and Barre HSAs	**WHIO data:** October 1, 2007 – September 30, 2009, with a three month run-out **Medicare:** 2005-2007 (20 percent sample)

	PCIP	Blueprint	WHIO
Patient attribution methods	Dartmouth's ACO patient assignment methodology (Patients must have one year of claims data to be eligible for assignment.)	Dartmouth's ACO patient attribution methodology (Patients must have one year of claims data to be eligible for assignment.)	N/A: population-based assessment
Level of analysis	Intervention vs. non-intervention	Intervention vs. non-intervention	County-level
Pools for comparison	*Intervention group* – Patients attributed to PCIP practices throughout the New York City area. *Comparison group* – Patients attributed to non-PCIP providers in New York City boroughs	*Intervention group* – Patients attributed to specific practices in St. Johnsbury, Burlington, and Barre *Comparison group* – Onpoint used propensity score matched cohorts for the non-Medicare data and Dartmouth used weighted counties.	N/A: statewide county-level assessment
Risk adjustment	Tiered risk adjustment was used for both Medicare and Medicaid populations: 1. Direct adjustment for age, sex, race (ASR) 2. ASR + income 3. ASR + income + prior year Hierarchical Condition Category (HCC) scores For the Medicaid population, whether a beneficiary was covered under managed care (at any time during the observation year) was also included in the third tier of risk adjustment.	Dartmouth tiered risk adjustment for Medicare: 1. Direct adjustment for age and sex (AS) 2. AS + income 3. AS + income + prior year HCC scores Onpoint used ERGs for risk adjustment for the commercial data.	Dartmouth tiered risk adjustment for Medicare: 1. Direct adjustment for age, sex, race (ASR) 2. ASR + income 3. ASR + income + prior year HCC scores Ingenix used ERGs as well as age and sex risk adjusters for non-Medicare data
Measure specifications	Dartmouth produced measures specifications for cost and utilization measures (see Appendix A1 and A2). Dartmouth used AF4Q measure specifications for the diabetes measures which are similar to HEDIS measures specifications (see Appendix A5).	Dartmouth produced measure specifications for cost and utilization measures (see Appendix A1 and A2). Dartmouth used AF4Q measure specifications for the diabetes measures which are similar to HEDIS measures specifications (see Appendix A5). Onpoint produced cost and utilization measure specifications for the commercial data (see Appendix A3). Onpoint strictly followed HEDIS measure specifications for the diabetes measures.	Dartmouth produced measure specifications for cost and utilization measures for the Medicare population (see Appendix A1 and A2). Ingenix produced measure specifications for the Non-Medicare population (see Appendix A4). AF4Q measure specifications were used for the diabetes measures which are similar to HEDIS measures specifications (see Appendix A5).

Identification of Multi-Payer Data Sources for Each Site

Health care utilization varies significantly across the U.S. health care system. The regional variation in spending and quality has tremendous implications for the development of health reform. How individuals use health care services is essential to crafting strategies for quality and cost improvement. Those utilization patterns can be better explained through analysis of provider payment systems. However, since utilization patterns are disparate system-wide, a multi-payer perspective is critical to capturing the full range, and use, of patient services.

The first step in this study was to identify multi-payer data sources for each site—Medicare and at least one other payer—to develop a multi-payer baseline. Dartmouth has collected Medicare data for each site detailed in the study. Medicare data from 2005 to 2007 were used to establish baselines for the Medicare population in each of the three sites. This database provided access to all physician claims for a random 20 percent sample of Medicare beneficiaries in the New York City area; Vermont's St. Johnsbury, Burlington, and Barre HSAs; and for Wisconsin.

In contrast to the Medicare data collection, the non-Medicare data was created differently in each site as a function of the population each community was targeting. For New York, Medicaid claims data were used from calendar year 2007. Medicaid claims were more relevant for analysis than commercial data because PCIP is designed to serve underserved populations like those traditionally benefiting from Medicare and Medicaid services. PCIP Medicaid data was acquired and managed by Tod Mijanovich at New York University (NYU).

WHIO used Medicaid and commercial insurance claims data (including Medicare managed care claims data) to develop an all-payer baseline that would identify areas of treatment and cost variation across the state. While Cigna and Aetna data and a portion of administrative services only (ASO) self-funded claims data are missing from the WHIO dataset, this should not have a significant impact on this study because these are not major payers in Wisconsin.[9] For the non-Medicare population, data were used from October 1, 2007, to September 30, 2009, with a three-month run-out. Non-Medicare data was analyzed by Ingenix, which also manages the data.

Onpoint Health Data analyzed the Vermont commercial data using insurance claims data for two HSAs in Vermont. Baseline data for the St. Johnsbury site contained commercial data for the period of July 1, 2007, to June 30, 2008. In Burlington, the commercial data was taken from October 1, 2007, to September 30, 2008. Vermont data includes eligibility and claims data from the Vermont Healthcare Claims Uniform Reporting and Evaluation System (VHCURES). Medicaid data is not yet included in the VHCURES database, but Vermont plans to use this in future evaluations of Blueprint. The Onpoint data for the non-Blueprint commercial population consists of non-Blueprint persons randomly drawn from the Vermont MVP plan, CIGNA, and Blue Cross populations in Vermont. This commercial data, however, excluded the disabled Medicare population and the largest health plan in one of the intervention sites in Vermont. The exclusion of data was considered in the analysis of study findings.

Interventions were initially launched in New York and Vermont in 2008 and were rolled out to different practice sites across multiple months and years. As a result, a method had to be developed for dealing

[9] ASOs are a significant part of the WI market (potentially as much as 60% of the market) and at present WHIO contains roughly 50% of the potential ASO block. This is sufficient for analyses at the state and county level, but does have implications at the local market and individual physician level.

with these staggered start-up dates. Dartmouth and NYU took a different approach to Onpoint's methods for the Vermont commercial population. For both Medicare and Medicaid in PCIP and for Medicare in Vermont, HSAs with start-up dates on or after October 1, 2008, were included as part of the pre-intervention baseline. Sites whose intervention began prior to October 1, 2008, were excluded and therefore utilization data patients attributed to them were not used in the analysis. The impact of this method is likely to be minimal in terms of data loss since very few PCIP and Blueprint practices had start-up dates prior to October 1, 2008. This approach was feasible for Medicare and Medicaid analysis because adequate amounts of data dating back well before October 2008 were available to construct a pre-intervention baseline.

While the Vermont Medicare data for St. Johnsbury, Burlington, and Barre HSAs were combined to ensure sufficient sample size (Dartmouth had access only to a 20% sample), Vermont commercial data were broken down by HSA because sample size was not a concern. However, Onpoint had access only to data dating back to 2007, which was a concern given Blueprint's 2008 launch. To ensure adequate data availability, Onpoint utilized a staggered start-up date approach whereby the pre-intervention period for each HSA consisted of a rolling 12 months of data prior to each sites actual state-up date. St. Johnsbury initiated its program on July 1, 2008, therefore, the pre-intervention period consisted of commercial data for St. Johnsbury from July 1, 2007, through June 30, 2008. By contrast, Burlington initiated its program on October 1, 2008, therefore, the pre-intervention period consists of commercial data for Burlington from October 1, 2007, through September 30, 2008. Barre, the third intervention site for Vermont, began in January of 2010 and is only included in the Vermont Medicare dataset.

Identification and Prioritization of Measures

Consistently defined comprehensive performance measures are crucial in evaluating the value of health care patients receive. Examining quality alone, or cost and utilization in isolation, would not provide an accurate assessment of value because high-cost care does not always imply quality and high-quality care does not always have to be expensive,[10,11,12] so a combination of cost, utilization, and quality metrics were selected for this project.

To select the measures, this project leveraged those measures that were already in the process of being developed for other initiatives underway in the three communities in order to facilitate greater alignment across initiatives. Some sites also chose to prioritize measures that were nationally endorsed. Second, where communities identified high-priority measurement gaps, this project aimed to fill them. Due to variation in the objectives and target populations of site interventions, some measures varied between sites while other measures were high-priorities for multiple sites. The development of a total cost of care metric, for example, was highly prioritized by all three communities.

Additionally, as part of the assessment process for the three sites, a series of community reports were developed using Medicare data to report on utilization, expenditure, and HEDIS measures at the regional level. These measures benchmarked surrounding areas and the U.S. national average. Through these reports, sites were able to see how they fared compared to local counties and major cities with

[10] Fisher, Elliott S., and others. (2003). "The Implications of Regional Variations in Medicare Spending. Part 1: The Content, Quality, and Accessibility of Care" *Annals of Internal Medicine* 138 (4), 273–287.
[11] Baicker, Katherine and Amitab Chandra. (2004). "Medicare Spending, the Physician Workforce, and Beneficiaries' Quality of Care" *Health Affairs.*
[12] Sirovich, Brenda, and others. (2006). "Regional Variations in Health Care Intensity and Physician Perceptions of Quality of Care," *Annals of Internal Medicine*, vol. 144, pp. 641–649.

similar geographic and population characteristics. The reports helped the sites identify measures that might be prioritized in developing a template for their evaluation process. The community reports can be found in Appendix D.

In some instances, the inability to access data or a lack of funding made it difficult to run measures consistently across all three sites. The ultimate selection of measures for each site was affected by site priorities, measures already being run by the sites and resource and data limitations. This section describes the measures that were selected for this project and explains any variation between sites.

Cost

Measuring the cost of care being provided allows sites to monitor the primary drivers of increasing health care costs and is necessary to analyze the value of the care patients are receiving. The primary cost measurements used in the project were total expenditures (per person per month), emergency department spending (dollars per 100 per month), and acute care hospital payments (per person per month). The total cost measure was broken down further into three separate measures: physician and hospital resource use per person-month; physician utilization (using standard prices) per person-month; and hospital utilization (using standard prices) per person-month.

The total expenditures measure was used across all payers for each site. The measure identifies the rate of payments per person-month for all facilities, providers and services covered under Medicare Parts A and B for the Medicare population. The total expenditures are the sum of positive Centers for Medicare and Medicaid Services (CMS), primary payer, and out-of-pocket payments. The measure was broken down into acute care hospital; durable medical equipment; evaluation and management; home health; hospice; imaging; long-term care; procedures; skilled nursing facilities; and testing.

A major difference in the cost measures used by the sites was around the decision to use allowed charges or standard prices, driven in large part by whether such data were available to each data source. Allowed charges represent the total amount reimbursement health plans have negotiated with care providers for given services and are often referred to as “agreed amounts.” Allowed charges have the advantage that they measure total reimbursement, including the amount patients are responsible for. They also have the advantage of assessing actual cost of care. However, allowed charges are also likely to vary depending on where a patient seeks care, making equivalent comparisons between institutions or regions very difficult. For example, if a person seeks care at a facility that only accepts insured patients and does not have a residency program, the allowed charges are likely to be lower than if they had gone to an academic medical center where teaching costs increased the rates of payments.

By contrast, standard prices attach a specific pricing schedule to utilization rates to remove these types of price variations. While standard prices have the advantage of enabling applies-to-applies comparisons across geographic areas and institutions and assessing the cost implications of utilization, they do not provide information on actual cost of care.

Dartmouth ran two separate cost measures for the Medicare populations in each community: one for standard prices and one for allowed charges for the Medicare data (allowed charges information are publicly available for Medicare). Because allowed charges data are also available in Medicaid data, NYU also ran both cost measures. To create standard prices, Dartmouth adjusted all standard prices nationally so that the sum of standard payments equaled the sum of actual CMS payments (not the sum of allowed charges).

Vermont prioritized allowed charges over standard prices when measuring cost of care. Information on allowed charges is often not publically available but is mandatory for payers to submit in Vermont. Additionally, Onpoint was not able to break the cost measure down by hospital utilization because they could not evaluate the diagnosis-related groups (DRGs) without additional funding, which limited the ability to create standard prices. In lieu of hospital expenditures, Onpoint was able to use relative value units (RVUs), as a proxy indicator of costs.[13]

In the case of WHIO, cost for the non-Medicare data was measured using standard prices with a standard unit pricing methodology developed by Ingenix. The Ingenix measure includes pharmacy data, whereas the Dartmouth measure specification does not include Part D prescription data.

In addition to measuring total cost, total emergency department spending was calculated across sites for the Medicare populations and the PCIP Medicaid population. The high cost of care in an ED setting meant this was an important measure for sites to monitor closely. This measure looks specifically at expenditures incurred on the day of an ED visit that did not result in admission to the hospital. For the Medicare population, the measure was broken down by allowed charges and CMS payments. Total cost for acute care hospital visits were also calculated for all payers and sites.

Due to resource constraints, Onpoint was only able to report emergency department visit expenditures from hospital outpatient billings. In Wisconsin, this measure was not identified as a first tier priority and was therefore not run in the WHIO non-Medicare populations.

Utilization

When analyzing the value of care being provided, it is important to also consider utilization. Differences in the volume of care can lead to variation in spending. A variety of measures were used in this project to identify rates of utilization: outpatient visits (including a breakdown of visits to primary care providers and specialists), hospital admission (including separate measures for surgery and other medical conditions), and length of stay measured in hospital days per 1,000 and ED visits per 1,000). These utilization measures were chosen for their ability to analyze the use and allocation of provider and other health care resources, free of any price differences that existed between sites. Consequently, this enabled a closer examination of the rates of events and the use of resources independent of spending.

Due to resource and data limitations, some variation existed between the utilization measures run across sites. All utilization measures were run for the PCIP Medicare populations and non-PCIP populations. ED hospital days were not run for the PCIP Medicaid population as Medicaid data does not include ED encounters that result in hospitalization. WHIO prioritized hospital admissions over the other utilization measures for the Medicare and non-Medicare populations. All utilization measures were run for the Vermont Medicare population. For the Vermont commercial population, outpatient visits could not be broken down by provider because resources were not available to link providers which would have resulted in limited accuracy in identifying provider specialties.

[13] Relative value units (RVUs) are used to determine allowable payment rates by applying a conversion factor to a measure of physician productivity. Johnson, Sarah E. and Warren Newton. *Resource-based Relative Value Unites: A Primer for Academic Family Physicians*. http://www.stfm.org/fmhub/fm2002/mar02/sa1.pdf and National Health Policy Forum. *The Basics: Relative Value Units*. 12 Feb. 2009. http://www.nhpf.org/library/the-basics/Basics_RVUs_02-12-09.pdf

Quality

HEDIS Diabetes Measures

Since claims data were the only data available to the sites, the ability to measure care quality at each site was limited. Four HEDIS diabetes metrics were used in this analysis (receipt of A1c test, receipt of eye exam, receipt of lipid test, and a composite measure of receipt of all three tests). Diabetes is one of the leading causes of death and disability in the United States. According to recent reports by the Center for Disease Control and Prevention (CDC), approximately 24 million people in the United States have diabetes and 57 million are at high risk for the disease. Diabetes puts people more at risk for kidney failure, cardiovascular diseases and contracting and acute illness resulting in worse health outcomes than non-diabetics. Timely screening for diabetic patients can significantly reduce the onset of complications for diabetics.[14]

Though these measures nominally assess treatment quality for diabetic populations, they were included in this project as proxies of overall quality of care, similar to the approach used by Baicker and Chandra in 2004. These relatively inexpensive and evidence-based procedures are indicative of effective care, which translates to quality of care.[15]

Agency for Healthcare Research and Quality (AHRQ) uses this diabetes composite measure as a key national health care quality indicator, and its use in this project would allow the communities to benchmark their performance against national averages moving forward.[16] The diabetes test measures, including the composite measure, were run across each of the sites, the only exception being the PCIP Medicaid population. New York Medicaid data is not broken down for HbA1c and lipid testing, so only the eye exam measure was used for this site.

Ambulatory Sensitive Condition Measures

Additionally, several ambulatory sensitive condition (ASC) measures (hospital admissions for ambulatory sensitive conditions per 1,000 person-months, number of inpatient days per ASC, and ED visits for ASCs per 1,000 person-months) were used as quality metrics in the study.

Hospital use accounts for approximately 31 percent of total health care expenditures according to a 2007 report to Congress by the Medicare Payment Advisory Commission. The report also found that 18 percent of Medicare hospital admissions result in readmission within 30 days, accounting for roughly $15 billion in spending, $12 billion of which was on potentially preventable readmissions. Similarly, when a patient enters the ED with a condition that is considered ambulatory care sensitive, there is a high likelihood that the ED visit could have probably been avoided.

The ASC-related measures used in this project identifies ED visits and hospital admissions for conditions (see Appendix A6) for which appropriate outpatient care could have potentially avoided the need for ED use or hospitalization. While it is difficult to know exactly how many admissions for ASC conditions are

[14] Centers for Disease Control and Prevention. (2011). National diabetes fact sheet: national estimates and general information on diabetes and prediabetes in the United States, 2011. Atlanta, GA.

[15] Baicker, Katherine and Amitabh Chandra. *Medicare Spending, Physician Workforce, And Beneficiaries' Quality of Care*. April 2004. Health Affairs.

[16] Agency for Healthcare Research and Quality. (2010). "National Healthcare Quality Report 2009" AHRQ Publication No. 10-003

preventable, these measures have been commonly used to assess patient access to care and the quality of care delivered in an ambulatory setting.[17]

Hospital admissions and ED visits for ambulatory sensitive conditions were measured across all sites with the exception of ED visits for ASC in the WHIO non-Medicare population.[18] Since these measures have not been endorsed by HEDIS, and funding was a factor in prioritizing measures for the non-Medicare WHIO data, this was not selected as a first priority measure.

Exhibit 3 provides the full list of the specific measures used in the study and the sites that utilized each measure in the analysis.

Exhibit 3. Measures Crosswalk by Measure Type, Site, and Payer[19]

		Wisconsin				Vermont		NYC	
		Medicare	Medicare Advantage	Commercial Plans	Medicaid	Medicare	Commercial Plans	Medicare	Medicaid
Cost	Measure 1: Total expenditures	X	X	X	X	X	X	X	X
	Measure 1a: MD and hospital resource use	X	X	X	X	X		X	X
	Measure 1b: Physician utilization (using standard prices)	X	X	X	X			X	X[20]
	Measure 1c: Hospital utilization (using standard prices)	X	X	X	X		X[21]	X	X[22]
	Measure 2: ED spending	X				X		X	X
	Measure 3: Acute care hospital payments	X	X	X	X	X	X	X	X
Utilization	Measure 4: Total outpatient visits					X	X	X	X
	Measure 5: Outpatient visits to primary care providers					X		X	X
	Measure 6: Outpatient visits to specialist physicians					X		X	X
	Measure 7: Hospital admissions	X	X	X	X	X	X	X	X
	Measure 8: Hospital admissions for surgery	X	X	X	X	X		X	X
	Measure 9: Hospital admissions for other medical conditions	X	X	X	X	X		X	X
	Measure 10: Hospital days					X	X	X	X
	Measure 11: ED visits					X	X	X	X
Quality	Measure 12: Hospital admissions for ambulatory-sensitive conditions	X	X	X	X	X	X	X	X

[17] Fitch, Kathryn and Kosuke Iwasaki. (January 2009). *Ambulatory-care-sensitive admission rates: A key metric in evaluating health plan medical-management effectiveness.* Milliman Consulting. Retrieved December 13, 2011, from http://www.nybgh.org/pdfs/ambulatorycare.pdf.

[18] The Onpoint method of identifying potentially avoidable Emergency Department visits varies from Dartmouth's and has been used for projects in Maine, New Hampshire, and Vermont. These diagnoses codes for outpatient Emergency Department visits represent approximately 22% of all outpatient Emergency Department visits – 30% in the Medicaid population, 22% in Commercial and Uninsured populations, and about 16% in Medicare populations. The alternative method proposed by Dartmouth may yield a very small proportion of Emergency Department visits in Medicaid and Commercial populations. This is an important measure to consider carefully, since for the Commercial and Medicaid populations, Blueprint and interventions may have a significant impact on the results.

[19] All measure specifications are included in the appendix.

[20] For standard payments, NYU used DRGs on the inpatient side and RVUs on the outpatient side to calculate payments.

[21] Onpoint used RVUs as a proxy for this measure.

[22] For standard payments, NYU used DRGs on the inpatient side and RVUs on the outpatient side to calculate payments.

	Measure 13: ED visits for ambulatory-sensitive conditions	X				X	X	X	X
	Measure 14a: HEDIS diabetes HbA1c	X	X	X	X	X	X	X	
	Measure 14b: HEDIS diabetes eye exam	X	X	X	X	X	X	X	X
	Measure 14c: HEDIS diabetes lipid test	X	X	X	X	X	X	X	
	Measure 14d: HEDIS diabetes composite measure (receipt of HbA1c, eye exam, and lipid test)	X	X	X	X	X	X	X	

Standardizing Measure Specifications for Selected Measures across Data Sources

After determining which measures mattered most to each community, the next step was to create a system where sites were able to implement the measures consistently across payers. An example of standardizing the measures is in site-specific lists of eligible ASC conditions. While these measures were measured per 1,000 person-months across all sites and payers, each site differed slightly in terms what conditions they included as part of their ASC definition (see Appendix A6).

In Wisconsin, Ingenix strictly followed the AHRQ list of eligible conditions, widely regarded as a national standard. Onpoint used specific conditions from AHRQ's "Using Administrative Data to Monitor Access, Identify Disparities, and Assess Performance of the Safety Net" report.[23] Dartmouth selected conditions that were specific to the Medicare population. For the Medicare data, the following conditions were considered ASC: angina; asthma; cellulitis; chronic obstructive pulmonary disease (COPD); congestive heart failure (CHF); convulsions; dehydration; diabetes; gastroenteritis; hypertension; kidney/urinary tract infection; and pneumonia over the total months of enrollment. Specific co-occurring surgical procedures were excluded from the measure and conditions such as low birth weight, which are not relevant for the Medicare population, were also excluded.

Site-specific support strategies were developed through input from each site. For example, while WHIO had an adequate data infrastructure in place, it had yet to implement specific payment and delivery reform initiatives at the time of this project. Therefore, this project aimed to help WHIO identify areas to target in reform efforts and develop robust, feasible methodologies for measuring impacts on cost and quality once those reform approaches are in place. In the case of Vermont and PCIP, which already have value-based reimbursement programs in place, guidance was provided to refine evaluation approaches and cross-site comparisons. Since each of the three sites already had data available, this project was able to focus on developing multi-payer evaluation baselines for assessing the impact of their health care reform initiatives moving forward. Given the years of data available, an evaluation of the impact of the interventions was not possible, but an infrastructure has been set in place for sites to do so in the future. Standardizing measure specifications across payers and sites still proved challenging. Limitations in resources and access to data along with population and payer differences led to some measure specifications being tailored to specific sites and payers.

Choosing a Level of Analysis and Developing Population Pools for Comparison

Because Vermont and New York have implemented actual interventions, while Wisconsin is still in the intervention planning process, the level of analysis needed to reflect the different study designs and is explained in further detail below.

[23] Billings, John. (October 2011). "Using Administrative Data to Monitor Access, Identify Disparities, and Assess Performance of the Safety Net." *AHRQ Archive-Home Page.* Retrieved December 13, 2011, from http://archive.ahrq.gov/data/safetynet/billings.htm.

Vermont and NYC

Because it was important to Vermont and New York to be able to illustrate the potential effects of their interventions, these two sites utilized a cohort study design that allowed for comparisons between treatment and comparison groups.

For Vermont, Blueprint wanted to assess impacts of patients whose physicians participated in its Blueprint medical home pilots, which began with physician groups located in three towns (Burlington, St. Johnsbury, and Barre). Due to data limitations, only Burlington and St. Johnsbury were included in the commercial data. For the Medicare populations, the three sites were consolidated and evaluated as a single unit as a result of sample size limitations (only a 20% sample was available for the Medicare data). For the commercial population, data was separated by HSA due to concerns around the intervention start-up dates. Blueprint launched in July 2008 for St. Johnsbury and October 2008 in Burlington Impacts were assessed for the Medicare population (ages 65–99 and disabled individuals ages 20 years and older) in addition to those with commercial insurance (non-disabled individuals ages 64 and younger).

For New York, the PCIP project wanted to assess impacts of patients whose physicians participated in TCNY's electronic medical records (EMR) intervention. The populations studied included both Medicaid (ages 20-64) and Medicare (ages 20–99 and disabled individuals younger than 65). It is important to note that the PCIP Medicaid data includes managed care beneficiaries. For this project, Medicaid recipients who have been considered managed care exempt,[24] and supplemental security income (SSI) recipients and those who are dually eligible for Medicare and Medicaid, have been excluded from the data to promote comparability between the treatment and comparison groups. Similar to the Vermont Medicare provider sample size issue, all PCIP providers were aggregated into one treatment group. Analysis was performed at the aggregate level rather than the individual physician or physician group level in order to ensure sufficient statistical power.

Analysts created comparison groups by attributing patients to specific practices for the intervention sites (patients attributed to PCIP practices throughout the New York City area and patients attributed to specific HSAs for Vermont).[25] For the Medicare data, Dartmouth assigned weights to controls such that the sum of control weights in each county equaled the number of cases, effectively controlling for regional variations in patient population. Since it is possible that people seeking care in a state are not actually residents of the area (e.g., "snowbirds" who may receive care from a physician in Vermont but officially reside in Florida), only those counties or boroughs where at least 5 percent of Medicare beneficiaries were assigned to a Blueprint or PCIP provider were included in the analysis. Comparison patients were selected from those same counties and boroughs.

In the case of NYC, this eliminated certain boroughs from being used for the Medicare populations, mainly the suburbs, due to having less than 5 percent of beneficiaries assigned. Once the comparison group was defined, Dartmouth weighted the group so that the sum of the weighted comparison group equaled the number of intervention patients in each county. For example, if one county had 20 intervention patients and 200 comparison patients, each comparison patient received a weight of 1/10.

[24] Patients are considered managed care exempt for HIV/AIDS, serious mental illness (SMI/SPMI) or OMRDD).

[25] Vermont's Blueprint preferred that Dartmouth use patient-level matching. However, sample size limitations resulted in Dartmouth using a matched cohort design (patient-level sampling could result in a smaller sample size. Since Vermont already presented a limited sample frame, this potential exclusion of controls would have weakened the power of the analysis).

As a result, the boroughs included in the analysis varied by payer data. In the case of the Medicare data, only Queens, the Bronx and Manhattan were used for the study, whereas in the case of Medicaid, all boroughs were included except for Staten Island.

Onpoint utilized propensity matching for attaining the comparison population of non-Medicare patients in Vermont. For Burlington, all matched cohorts were residents of the intervention site. However, for St. Johnsbury, residents of Morrisville and Newport were pooled into the comparison group because of insufficient participant numbers in St. Johnsbury alone. Using propensity score matching, participants in intervention sites were each matched with four comparison patients. Analysis was performed independently for each site due to the different start-up dates.

Though expectations for observable effects of the interventions are slight because of the recent date of intervention, to compensate for the temporal challenges and improve the power of statistical analysis, annual rates from the cohorts were also calculated and combined to produce a single (time-adjusted) rate. This provided a time-adjusted difference to examine between cases and controls of the Medicare population.

Wisconsin

In order to accurately assess potential targets for future reform, WHIO needed a statewide view on cost and quality variations, so non-Wisconsin residents were excluded from the data. Because WHIO was not evaluating an intervention but looking across the state in order to identify cost and quality variations that may be targets for future health and payment reform efforts, the analysis in Wisconsin was performed based on county of residence. Hospital service areas were considered potential units of analysis. However, using HSAs exposed the data to possibly reflect patients and care in neighboring states. Therefore, counties were used as the unit of analysis to avoid incorporating data that could have included out-of-state residents.

Selecting and Implementing a Consistent Patient Attribution Method

In order to allocate patients to either the comparison or treatment groups, it was necessary to first determine which physicians were participating in either the Vermont Blueprint or New York PCIP programs and then assign patients to those providers based on whom they made a plurality of their ambulatory visits to.

Eligible Providers

A list of all providers affiliated with the provider sites was provided with which to record the eligible providers in the study.

For situations where the number of individuals assigned to these excluded physicians comprised more than 5 percent of the commercial population, a corrective adjustment was performed where the previously excluded physicians were included using the same criteria used for individuals in the Medicare population. To ensure consistency in the results, the Medicare assignment was determined to be the default. Only physicians who remained unassigned using the Medicare method were assigned to hospital networks based on their levels of experience treating the commercial population.

Various methods are used to identify providers in claims data making it difficult to consistently identify providers. For example, because providers are categorized by specialty, identifying providers in the PCIP Medicaid data proved to be a challenge as provider specialty information was often either missing altogether or a provider was listed as having multiple specialties. If a provider's specialty is not given in

Medicaid claims, that provider is often assumed to be a primary care physician (PCP) even if that may not be the case. Furthermore, providers often identified themselves as primary care physicians even though they were actually practicing as specialists.

To address this issue, Dartmouth used several years of claims data and categorized providers by the most frequently billed specialty. Dartmouth and NYU were able to then use a national provider identification (NPI)-unique provider identification number (UPIN) crosswalk to match some providers for the PCIP Medicaid and Medicare populations. Of 2,500 providers operating in the NYC region, only 1,500 had NPIs and 900 had UPINs. Of those providers, only 550 provided care to at least one Medicare beneficiary during the pre-intervention period. Furthermore, Dartmouth was unable to link UPINs to NPIs for those providers billing primarily as OB-GYNs and pediatricians for PCIP Medicaid, as those specialists typically do not participate in the Medicare program and many of the providers typically do not bill for outpatient visits. As a result, roughly one-quarter of the providers were ultimately included in the analysis.

For the PCIP Medicaid data, NYU was able to match ancillary fields from provider files with New York State license numbers. While this method was especially time consuming, it did allow for a much higher rate of successful matches.

Attributing Patients

Once the participating providers were identified for Vermont and New York, a method had to be selected to consistently attribute patients to those providers. The patient attribution methodology developed by Dartmouth was used for this project. This methodology can be used by commercial payers and was used in the Medicare and Medicaid populations to create patient pools for each provider that were aggregated into the intervention and comparison group.

The process began when individuals were assigned to a provider based on where they received the plurality of their non-inpatient evaluation and management (E&M) visits during the measurement year. Individuals who had no outpatient E&M visits in the study period were excluded from the analysis. In order to be attributed to a physician, patients who survive through the end of the measurement year must have been enrolled with the carrier for at least nine cumulative months of the calendar year; patients who died during this year need only to have been enrolled for one month. These patients were then attributed to a single provider and reassigned annually. A flow diagram describing this patient attribution methodology is provided in Exhibit 4.

The lack of commonalities between Medicare, Medicaid, and commercial data made it difficult to fully standardize patient attribution methods across all sites and payers and required minor adjustments to the procedure described above. For example, in PCIP and Vermont, all providers were pooled together, so patients were either assigned to an intervention group site or a non-intervention group individual physician. In Vermont, the comparison group for the Burlington region may be skewed because there is a possibility that a majority of local residents receive care from physicians associated with the Blueprint intervention sites.

Selecting and Implementing Consistent Risk Adjustment Methods

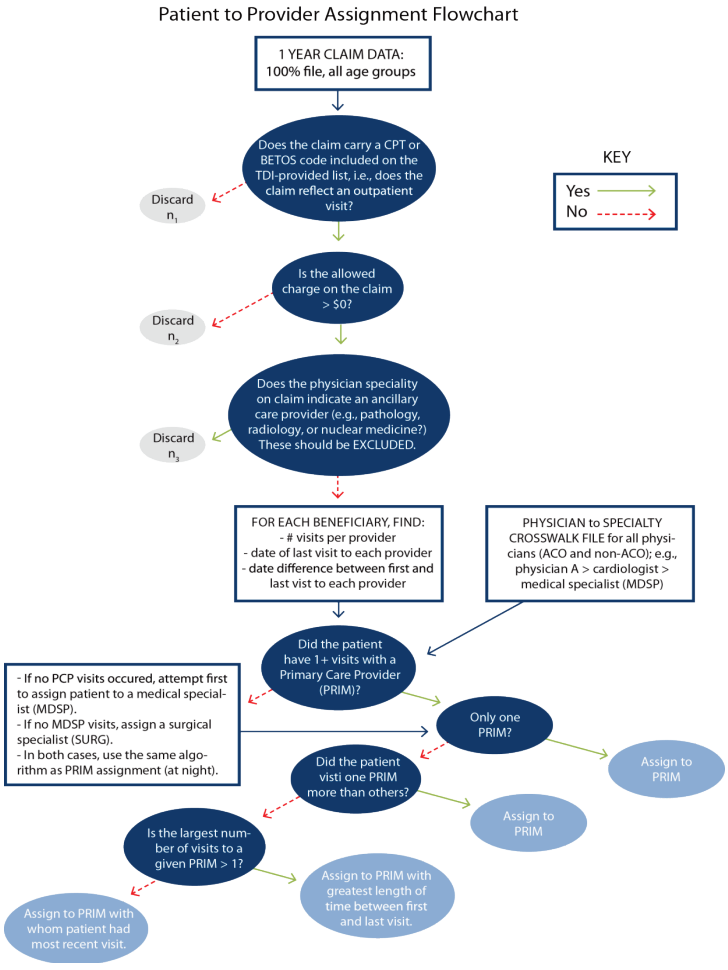
Levels of health care utilization, spending, and health outcomes are influenced by a multitude of factors and include the quality of the care physicians deliver but also include factors like age, sex, and underlying health status. In order to properly account for possible mediating effects on health care utilization and spending, this project employed standard direct adjustment techniques to risk-adjust cost

and utilization measures. HEDIS quality measures, however, were not risk-adjusted as measures pertaining to process measures (e.g., receipt of guideline-based recommended care) are generally not risk-adjusted as they should not vary as a function of demographic factors. Tiers of risk adjustment allow for various levels of adjustment to be used so long as they are statistically significant. When results are consistent across each tier, the results are unambiguous.

Dartmouth included multiple risk adjustment tiers for the Medicare population analyses in order to account for the possible mediating factors presented by demographic variables, including factors like sex, education, and chronic conditions that could confound the study analysis if patients differentially selected the providers of their care. The three tiers of risk-adjustment were the following:

1. Direct adjustment for age, sex, and race (ASR)
2. Adjustment for ASR + zip education + zip income
3. Adjustment for ASR + zip education + zip income + prior year hierarchical condition category (HCC) scores + presence/absence of a select list of chronic conditions[26]

Exhibit 4. The Dartmouth Institute's Patient Attribution Methodology



[26] For list of conditions, see Appendix A.6

The first tier represents a preliminary risk adjustment method. Since it is a basic risk adjustment, tier two and three are used to provide substantive further risk adjustment where possible. For tier two risk adjustment, income and education data were pulled from Census data at the zip-code level. Building on tier two, tier three incorporated HCC scores. HCC scores are designed to predict average health care expenditures using demographics and comorbid conditions/events as outlined by CMS.

Only age and sex adjusters were used as a first tier of risk adjustment for the Vermont commercial population as VHCURES does not include data on race. Dartmouth also excluded race in adjusting for the Vermont Medicare population in order to be consistent across payers and because a small percentage of Vermont's Medicare population is black. Because the comparison and intervention populations used for Vermont were mostly white, not adjusting for race should have little impact on results. Onpoint used episode risk group (ERG) scores compiled by Ingenix software from claims data for a more sophisticated risk adjustment method for Vermont commercial data. These scores were used in addition to age and sex, chronic conditions, and payments.

For WHIO, episode treatment groups (ETG) represented three units of analysis for clinical comparison: episodic (acute/chronic), non-episodic (preventive), and ungroupable (invalid codes). Each group was partitioned into severity levels if such variation was observed within the episode. Where variation existed within the episode, a score of 1 to 4 was set by age, gender, and measurement of comorbidities. While the Dartmouth methods adjust risk at the person level, Ingenix's risk adjusters adjust data at the episode-level.

Finally, for the Medicaid data, PCIP risk adjustment pertained to ASR. In the third tier of risk adjustment, managed care enrollment was also taken into account.

Limitations

The study is subject to several common research limitations. The first limitation is that physician tracking proves problematic for a multi-year, cross-sectional design. Many physicians working in intervention areas during baseline data collection are potentially removed during follow-up periods. The change in physicians practicing after implementation of the intervention could alter the pre- and post-intervention data. Therefore, some post-intervention data results could be attributed to a shift in the physician profile of the area.

A second limitation is the relatively small sample size. To thoroughly analyze intervention-specific effects, a larger sample size may be necessary. In light of these considerations, the Medicare populations in the three Blueprint intervention sites were pooled into one intervention group. Nevertheless, evaluating sub-populations of a sample that is already relatively small (e.g., the diabetic population in the three Blueprint HSAs, broken further into comparison and intervention groups) posed challenges and does not provide adequate statistical power to draw inferences about whether there were statistically significant differences between the quality and cost of care received by treatment and comparison groups. However, as Blueprint expands and the number of patients included in the treatment group grows, power should gradually increase. From a programmatic perspective policy officials administering Blueprint felt that having a baseline that enables them to track improvements over time, even if not sufficiently powered to draw statistically significant conclusions about the medical home pilot, would nevertheless be very useful information to have from a clinical and policy perspective.

Additionally, illness adjustment, whether via HCC, ETG, ERG or other established methods, may produce biased findings because the degree of diagnosis coding is highly correlated with spending and thus higher spending areas appear to have sicker populations.[27]

Finally, the lag in reporting time for Medicare data was another major limitation. When this project began, Medicare data from 2005–2007 was available, limiting any possibility of incorporating post-intervention data. Additionally, all baseline Medicare data is assessed using data from the years 2005 through 2007 precluding any possibility of using baseline data staggered around intervention start-up dates for the intervention sites.

[27] Welch, Gilbert H, et. al. (2011). Geographic Variation in Diagnosis Frequency and Risk of Death Among Medicare Beneficiaries. 305(11):11113-1118.

Chapter 3
Study Findings

Chapter 3: Study Findings

This section describes the baseline findings from the cross-payer analyses for PCIP, Blueprint, and WHIO. The findings are broken down by site, concluding with an assessment of the Medicare population across all three sites. Moving forward, these findings will provide Blueprint and PCIP with baseline data that will be critical for assessing the impacts of their interventions and can help identify priority areas of focus for the development of payment reform interventions in Wisconsin. A summary of key findings is provided at the end of the chapter.

Each site profile begins with a description of characteristics of the payer populations that were studied for each site. Because differences in demographic factors like socioeconomic status and health status can sometimes lead to differences in health outcomes, utilization, and cost independent of the effects of health interventions, it is important to first examine the populations in each site, particularly in New York and Vermont where the baseline results will be used in the future to evaluate the effectiveness of various payment and delivery reform models. The Blueprint and PCIP tables each break down patient characteristics by those attributed to the comparison and intervention populations. In the case of WHIO, analysis was conducted at the county-level, so county populations of interest are described using 2010 Census data.

Multi-payer baseline data are then assessed by identifying patterns in cost and utilization. For PCIP and Blueprint, relative risks are used to identify differences in spending and utilization between the intervention population and the comparison group. For WHIO, data are displayed in a county-level comparison. The impact of the tiers of risk adjustment applied to the cost and utilization data is also evaluated in this section. The data displayed for each site reflects the fullest level of risk adjustment run for each group.

While cost and utilization measures are critical to bending the health care cost curve, these measures alone have limited utility because they do not assess health care value, which is a measure of cost or utilization relative to quality of care or health outcome. This section also examines a select set of quality metrics relative to total cost of care to assess the value of care patients were receiving at baseline in both PCIP and Blueprint. For WHIO, this section also identifies consistently high or low-value counties across payers for a select set of quality metrics.

The study findings conclude with a cross-site comparison of the value of care Medicare patients are receiving in each of the sites, benchmarked against state and national averages.

Vermont Blueprint

Blueprint and Non-Blueprint Population Overview by Payer

Medicare

Concerns around sample size led to the pooling of intervention HSAs (Barre, Burlington, and St. Johnsbury) in Vermont for the Medicare analysis. As Exhibit 5 shows, demographic characteristics are similar for the intervention assigned population and the geographically selected comparison group for the Medicare population. These similarities indicate that, at baseline, the Medicare population utilizing Blueprint is representative of the people in the areas served. The exception to this was education level (as assessed through average percentage of adults with different levels of educational attainment living in patients' residential zip codes), for which 0.6 percent more of the Blueprint population had college degrees when compared to the zip education of non-Blueprint group ($p<.001$). The overall HCC score, a

linear predictor of spending, was similar between groups, but it should be noted that 20.6 percent of the Blueprint population had diabetes compared to 16.5percent of the non-Blueprint population ($p<0.05$). The Blueprint population also had lower rates of mortality with 1.5 percent fewer deaths ($p<.05$).

Exhibit 5: Blueprint Medicare Population Profile

Variable	Blueprint Population		Non-Blueprint Population		Diff
	N	%	N[1,2]	%	P (X^2)
Overall (weighted):	**1131**	**100**	**1131**	**100**	
Age:					
20–49	76	6.72	77	6.85	
50–64	100	8.84	92	8.11	
65–69	261	23.08	233	20.62	
70–74	216	19.10	231	20.40	
75–79	186	16.45	193	17.08	
80–99	292	25.82	305	26.95	0.73
Mortality:					
Survived 2007	1097	96.99	1080	95.51	
Died	34	3.01	51	4.49	0.03
Sex:					
Female	681	60.21	662	58.49	
Male	450	39.79	469	41.51	0.40
Race recorded in claims records:					
Black	-	-	-	-	
Hispanic	-	-	-	-	
Other or Missing/Unknown	1124	99.38	1126	99.55	0.24
Zip-Level Median Income[3]					
Lowest quintile	-	-	-	-	
2	350	30.95	341	30.15	
3	173	15.30	221	19.55	
4	303	26.79	286	25.32	
Highest quintile	278	24.58	247	21.82	0.06
Education: Average percent of adults living in patients' residential zip codes with:[4]					**p (t-test)**
Less than high school	8.56		8.69		0.17
High school diploma	65.32		65.82		0.002
College degree	26.12		25.50		0.0082
Co-morbidities:[5]					**p (t-test)**
Overall score (HCC):	0.81		0.82		0.47
Co-morbidities:[5]	**N**	**%**	**N**	**%**	**p (X^2)**
Disability	291	25.73	263	23.25	0.39
Diabetes	233	20.60	186	16.47	0.04
CHF	59	5.22	56	4.98	0.97
IVD	47	4.16	52	4.61	0.87
COPD	86	7.60	83	7.38	0.98
Dialysis/renal failure	21	1.86	33	2.93	0.24
Liver disease (severe)	-	-	-	-	-
Alcoholism	11	0.97	13	1.14	0.93
Hip fracture	11	0.97	12	1.05	0.98

1. Ns for non-HIT population ("comparison group") are weighted by county "case" contribution.
2. Cells <11 are not show in accordance with CMS suppression rules.
3. Zip-level median household incomes, US Census 2000 data.
4. Zip-level education statistics, US Census 2000 data.
5. Co-morbidity data missing for 2.47% of HIT and 1.80% of non-HIT populations.

Commercial Population[28]

For the commercial population, analysis was broken down by the two HSAs—St. Johnsbury and Burlington. As Exhibits 6 and 7 demonstrate, the comparison groups in both sites were similar overall to the intervention groups. Exceptions to this were higher rates of diabetes (7.6% for Blueprint and 5.9% for non-Blueprint, p <.01) and lower rates of asthma (1.9% for Blueprint and 2.9% for non-Blueprint, p<.01) in St. Johnsbury.

Differences in overall health status were measured using ERGs that predict health risks using episodes of care from claims data as well as demographic variables. A higher ERG score indicates a higher risk of health care costs or utilization. Differences in ERG scores approached significance (p=.06) in St. Johnsbury with the Blueprint group having slightly higher ERG scores than those not participating in Blueprint. In Burlington, differences in ERG scores were statistically significant (p=.04) with the Blueprint population having higher ERG scores overall. Blueprint participants also had 0.9 percent higher rates of coronary heart disease in Burlington (p<.01).

Exhibit 6: Blueprint Commercial Population Profile for St. Johnsbury[29, 30, 31]

Measure	Blueprint Participants — Rate (Count)	Potential Control Group — Rate (Count)	Diff. p(chi-square)	Matched Controls — Rate (Count)	Diff. p(chi-square)
Total	**100.0% (1,882)**	**100.0% (12,459)**		**100.0% (7,528)**	
Age (years)					
18–29	8.6% (162)	14.4% (1,788)		8.9% (667)	
30–39	15.5% (291)	17.7% (2,201)		14.4% (1,081)	
40–49	26.1% (491)	25.7% (3,198)		26.3% (1,977)	
50–59	35.3% (665)	30.9% (3,848)		35.7% (2,689)	
60–64	14.5% (273)	11.4% (1,424)	<.01	14.8% (1,114)	0.82
Gender					
Female	55.8% (1,050)	52.2% (6,500)		56.7% (4,269)	
Male	44.2% (832)	47.8% (5,959)	<.01	43.3% (3,259)	0.47
ERG Score Range					
0	13.4% (252)	20.6% (2,571)		15.5% (1,169)	

[28] New analyses currently underway with Blueprint will potentially have higher numbers of attributed members, tighter and more consistent attribution methods, selection of a control population that will have a tighter match by having a primary care physician visit at baseline, and the evaluation will be based on practice location.

[29] Due to a small number of possible matches in St. Johnsbury, the St. Johnsbury control group also includes participants from other nearby towns with similar demographics. The final control group includes 1,711 residents of St. Johnsbury, 3,246 residents of Morrisville and 2,571 Newport residents.

[30] *First Evaluation of Vermont Blueprint Using Commercial Claims Data—Methods. Onpoint Health Data: May 2011.*

[31] Payment category is measured in person-years

0.0001–04999	17.7% (334)	21.6% (2,693)		19.4% (1,463)	
0.5000–0.9999	18.7% (352)	17.3% (2,158)		17.4% (1,313)	
1.0000–1.9999	23.1% (434)	18.8% (2,341)		21.5% (1,618)	
2.0000–4.9999	20.9% (393)	16.9% (2,109)		20.1% (1,512)	
5.000 +	6.2% (117)	4.7% (587)	<.01	6.0% (453)	0.06
Payment Category (annual costs)					
0	9.8% (185)	17.2% (2,139)		10.0% (756)	
$1 – $999	29.7% (559)	33.8% (4,210)		30.1% (2,264)	
$1,000 – $9,999	48.4% (911)	40.7% (5,073)		48.6% (3,655)	
$10,000 – $39,999	10.3% (193)	7.3% (905)		9.7% (733)	
$40,000 +	1.8% (34)	1.1% (132)	<.01	1.6% (120)	0.91
Disease Prevalence					
Any Chronic Condition	29.0% (546)	20.3% (2,527)	<.01	28.0% (2,106)	0.37
Asthma	1.9% (36)	2.2% (271)	0.46	2.9% (222)	0.01
COPD	1.2% (23)	0.7% (93)	0.03	1.1% (83)	0.66
CHF	0.2% (3)	0.2% (30)	0.49	0.4% (27)	0.17
Coronary Heart Disease	2.9% (55)	1.6% (201)	<.01	2.3% (176)	0.14
Hypertension	15.4% (290)	10.3% (1,283)	<.01	14.4% (1,087)	0.29
Diabetes	7.6% (143)	4.2% (519)	<.01	5.9% (443)	<.01
Depression	5.8% (110)	4.5% (556)	<.01	5.8% (438)	0.96

Exhibits 6 and 7 display data on the intervention and control populations, where the *Blueprint participants'* represent the intervention population and the *matched controls* represent the control population.

The *potential control group* is the resident population of which the matched controls are a subset. Differences between the resident population and the matched controls demonstrate the impact that propensity score matching had on the control group. For example, prior to propensity score matching, the control group had statistically significant rates of chronic conditions (p<.01). Propensity score matching was able to eliminate these differences between the control and intervention populations (p=.37).

Exhibit 7: Blueprint Commercial Population Profile for Burlington[32]

Measure	Blueprint Participants — Rate (Count)	Potential Control Group — Rate (Count)	Diff. p(chi-square)	Matched Controls — Rate (Count)	Diff. p(chi-square)
Total	**100.0% (2,105)**	**100.0% (47,367)**		**100.0% (8,420)**	
Age (years)					
18–29	9.5% (200)	19.0% (8,990)		9.3% (780)	
30–39	12.6% (265)	19.1% (9,052)		12.7% (1,070)	
40–49	26.4% (556)	26.3% (12,460)		25.7% (2,166)	
50–59	36.9% (777)	27.1% (12,830)		37.4% (3,149)	
60–64	14.6% (307)	8.5% (4,035)	0.06	14.9% (1,255)	0.96
Gender					
Female	53.0% (1,115)	50.9% (24,089)		52.2% (4,422)	
Male	47.0% (990)	49.1% (23,278)	<.01	47.5% (3,998)	0.71

[32] *First Evaluation of Vermont Blueprint Using Commercial Claims Data—Methods. Onpoint Health Data: May 2011.*

ERG Score Range					
0	13.5% (285)	20.2% (9,561)		14.8% (1,245)	
0.0001–04999	21.4% (451)	24.1% (11,408)		23.8% (2,008)	
0.5000–0.9999	19.6% (413)	18.2% (8,633)		19.2% (1,620)	
1.0000–1.9999	20.4% (429)	18.2% (8,642)		19.5% (1,638)	
2.0000–4.9999	19.5% (410)	15.4% (7,302)		18.0% (1,516)	
5.000 +	5.6% (117)	3.8% (1,821)	<.01	4.7% (393)	0.04
Payment Category					
$0	9.9% (208)	15.9% (7,552)		10.7% (897)	
$1 – $999	29.9% (630)	32.7% (15,466)		30.7% (2,589)	
$1,000 – $9,999	50.7% (1,068)	43.9% (20,772)		50.1% (4,215)	
$10,000 – $39,999	8.4% (177)	6.7% (3,192)		7.7% (646)	
$40,000 +	1.0% (22)	0.8% (385)	<.01	0.9% (73)	0.52
Disease Prevalence					
Any Chronic Condition	19.8% (417)	17.5% (8,284)	<.01	18.8% (1,580)	0.27
Asthma	2.3% (48)	2.4% (1,141)	0.71	2.2% (182)	0.74
COPD	0.3% (06)	0.4% (170)	0.58	0.5% (41)	0.21
CHF	0.4% (9)	0.2% (80)	<.01	0.2% (18)	0.08
Coronary Heart Disease	2.4% (50)	1.2% (568)	<.01	1.5% (126)	<.01
Hypertension	8.6% (182)	8.0% (3,812)	0.32	9.2% (775)	0.43
Diabetes	4.3% (90)	3.1% (1,485)	<.01	3.8% (324)	0.37
Depression	4.7% (98)	4.7% (2,242)	0.87	4.1% (347)	0.28

Baseline Cost and Utilization Assessment by Payer

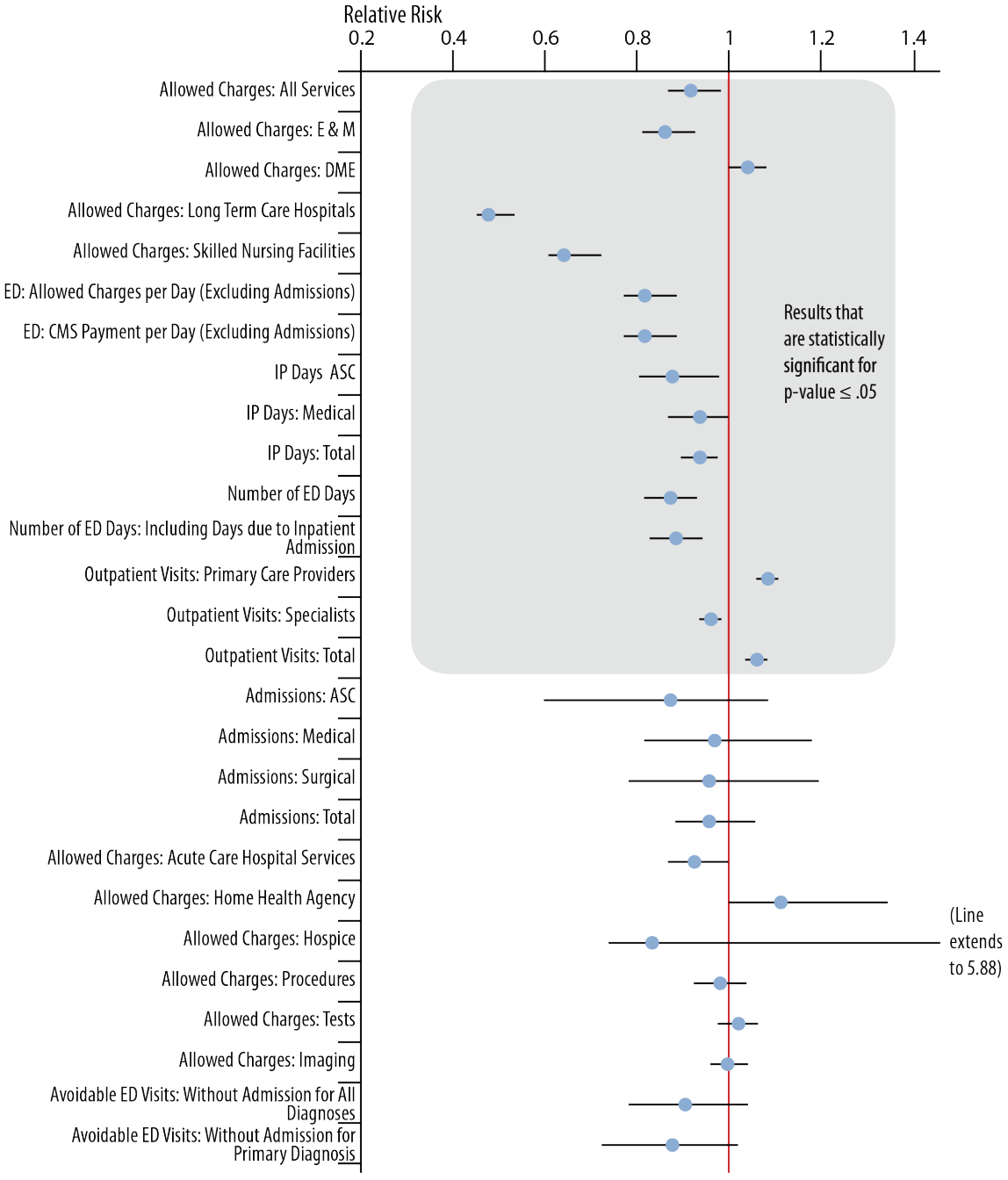
Medicare Population

Full results can be found in Appendix B1 and Exhibit 8 summarizes the cost and utilization data for the Blueprint Medicare population relative to the non-Blueprint Medicare population after risk adjustment.[33] Overall, Medicare patients in the Blueprint group had lower rates of cost and utilization at baseline. In three out of 15 statistically-significant results ($p \leq .05$), Blueprint patients were 3 to 12 percent more likely to have higher cost or utilization than the comparison group: allowed charges for E&M services; outpatient visits for primary care providers; and total outpatient visits. For 12 of the statistically-significant results, rates were lower in the Blueprint group. Utilization was 7 to 18 percent lower in the Blueprint population for measures around inpatient days and ED visits. Blueprint allowed charges and ED CMS payments were much lower than the non-Blueprint patients. Allowed charges for all services were 10 percent lower in the Blueprint group, with allowed charges for long-term care being 51 percent lower.

[33] Data has undergone level 3 risk-adjustment for VT: age, sex, education, income and HCC scores.

Exhibit 8: Relative Risk of Cost and Utilization Measures for Blueprint vs. non-Blueprint Medicare Population

Relative Risk of Cost and Utilization Measures for Blueprint and non-Blueprint Medicare Populations



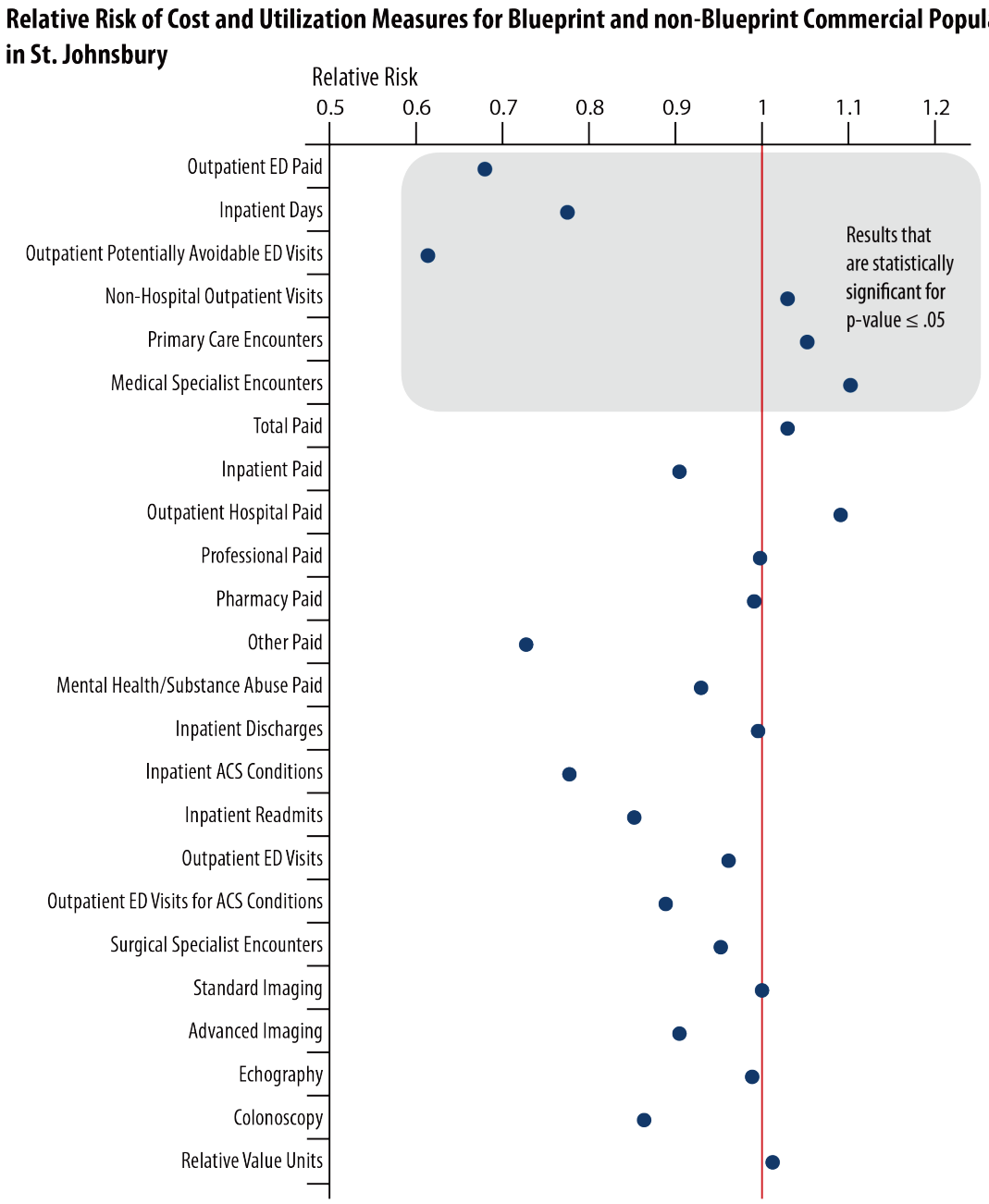
The relative risk table displays the ratio of something occurring in the intervention group relative to the comparison (non-Blueprint) group and therefore demonstrates differences in the baseline data between the two groups. For measures where the relative risk is greater than 1, the event happens more frequently or costs more in the Blueprint group. When the relative risk is lower than 1, the opposite is true and the event occurs or costs more in the non-Blueprint group. For example, a relative risk of 1.05 means that an event happens or costs 5% more in the Blueprint group.

When the confidence interval of a relative risk crosses 1, there were no statistically significant differences between the results for the intervention and comparison groups. The relative risks displayed above have been adjusted using tier 3 methods (AS + zip income + prior year HCC scores). The bars on each relative risk represent the 95% confidence interval.

Commercial Population

Exhibit 9 summarizes the cost and utilization findings for the Blueprint commercial population and the non-Blueprint commercial population in St. Johnsbury. The full set of results for the commercial St. Johnsbury population can be found in Appendix B7. Out of six statistically significant results ($p \leq .05$), Blueprint participants in the St. Johnsbury commercial population had higher rates of cost and utilization in half of those measures. St. Johnsbury Blueprint participants were 3 percent more likely to have higher rates of non-hospital outpatient visits, 5 percent more likely to have primary care encounters, and 9 percent more likely to have medical specialist encounters than those attributed to the comparison group in the commercial population. However, Blueprint commercial patients in St. Johnsbury were also 31 percent less likely to have outpatient ED expenditures, 39 percent less likely to have outpatient potentially avoidable ED visits, and 23 percent less likely to have inpatient days than those not participating in Blueprint.

Exhibit 9. Relative Risk of Cost and Utilization Measures for St. Johnsbury Blueprint vs. non-Blueprint Commercial Populations



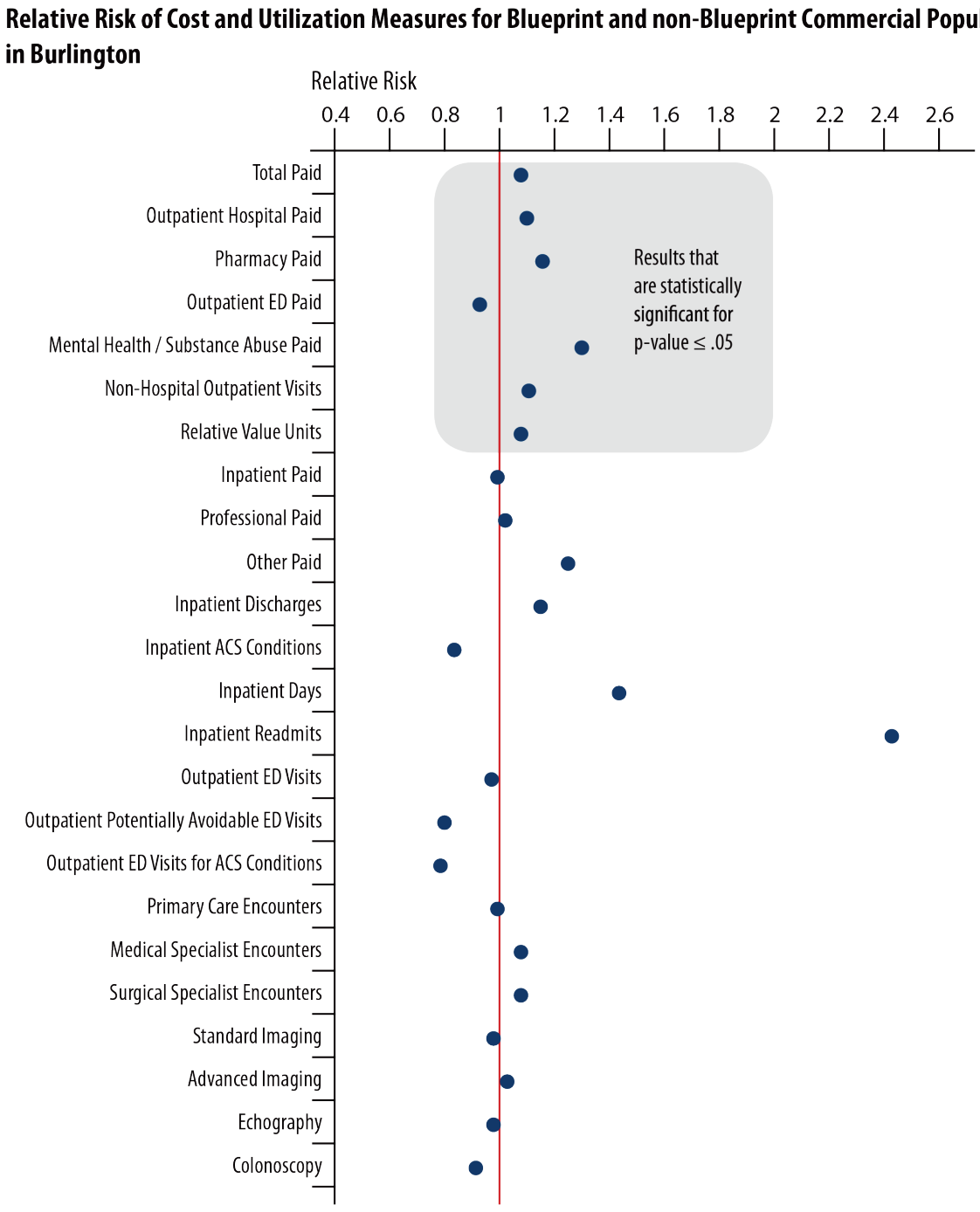
The relative risk table displays the ratio of something occurring in the intervention group relative to the comparison (non-Blueprint) group and therefore demonstrates differences in the baseline data between the two groups. For measures where the relative risk is greater than 1, the event happens more frequently or costs more in the Blueprint group. When the relative risk is lower than 1, the opposite is true and the event occurs or costs more in the non-Blueprint group. For example, a relative risk of 1.05 means that an event happens or costs 5% more in the Blueprint group.

The relative risks displayed above have been adjusted using ERGs.

Exhibit 10 displays the cost and utilization of Blueprint participants relative to non-Bluperint participants for the commercial population in Burlington. The full set of Burlington results can be found in Appendix B7. In six out of seven statistically significant results ($p \leq .05$), Burlington Blueprint participants had higher rates of cost and utilization than their non-Blueprint counterparts. Participants in Burlington's

commercial population had between 7 and 15 percent higher allowed charges for outpatient hospital outpatient payments, pharmacy payments, and total costs. Participants had 29 percent higher costs for mental health and substance abuse. Non-hospital outpatient visits were 11 percent higher and RVUs were 8 percent higher amongst patients in Burlington attributed to Blueprint. Outpatient ED expenditures was the only statistically significant measure for which Blueprint had a lower cost or utilization than non-Blueprint counterparts in the Burlington commercial population.

Exhibit 10. Relative risk of Cost and Utilization Measures for Burlington Blueprint vs. non-Blueprint Commercial Populations



The relative risk table displays the ratio of something occurring in the intervention group relative to the comparison (non-Blueprint) group and therefore demonstrates differences in the baseline data between the two groups. For measures where the relative risk is greater than 1, the event happens more frequently or costs more in the Blueprint group. When the relative risk is lower than 1, the opposite is true and the event occurs or costs more in the non-Blueprint group. For example, a relative risk of 1.05 means that an event happens or costs 5% more in the Blueprint group.

The relative risks displayed above have been adjusted using ERGs.

Impact of Risk Adjustment on Results by Payer

As noted earlier, the above analysis on baseline costs and utilization was conducted using the highest tier of risk adjustment in each payer population. Separate analysis was undertaken in order to assess the relative impact of different tiers of risk adjustment on both statistical significance as well as magnitude and direction of relative risk.

Out of 28 measures, risk adjustment methods made a difference in statistical significance in only four measures in the Blueprint Medicare results (Exhibit 11). For example, in the case of ASC admissions, level 1 risk adjustment (age + sex) resulted in a statistically significant difference between Blueprint and non-Blueprint populations, which disappeared using level 2 or level 3 methods. Risk-adjustment methods also did not affect directionality of risk nor did it substantially affect the magnitude of the measure. Because commercial data was risk adjusted using ERGs during the propensity score matching, there are not multiple tiers of risk adjustments.

Exhibit 11: Statistical Significance of Risk Adjustment Methods on Blueprint vs. Non-Blueprint Results for Medicare

		Non-Blueprint vs. Blueprint		
		Relative Risk d/		
		Level 1 [a/]	Level 2 [b/]	Level 3 [c/]
ASC Admissions	**Relative Risk**	**1.31**	**1.28**	**1.19**
	95% Confidence Interval	1.01 - 1.7	0.99 - 1.66	0.94 - 1.52
	p-value	0.04	0.06	0.16
Inpatient Days: Surgical	**Relative Risk**	**1.06**	**1.10**	**1.06**
	95% Confidence Interval	0.99 - 1.14	1.02 - 1.18	0.98 - 1.14
	p-value	0.11	0.01	0.13
Avoidable ED Visits Without Admission for Primary Diagnosis	**Relative Risk**	**1.20**	**1.23**	**1.20**
	95% Confidence Interval	0.99 - 1.45	1.01 - 1.49	0.99 - 1.45
	p-value	0.07	0.04	0.07
Allowed Charges: Home Health Agency	**Relative Risk**	**0.88**	**0.87**	**0.89**
	95% Confidence Interval	0.73 - 0.99	0.72 - 0.98	0.75 - 1
	p-value	0.03	0.02	0.05

[a/] The level 1 adjusters include Age (20-49, 50-64,65-69,70-74,75-79, and 80-99) and Gender (male and female).

[b/] The Level 2 adjusters include Education and Income factors in addition to all Level 1 adjusters. Zipcode rates of various levels of schooling (less than a high school diploma, high school diploma, and at least a college degree) are used to control for education. Medicaid status at the person-level as well as income quintile distributions at the zip code level is used to control for income.

[c/] Level 3 adjusters include the CMS Hierarchical Condition Category (HCC) risk scores in addition to the Level 1 and Level 2 adjusters. The relevant performance year is used to generate the HCC score.

The impact of risk adjustment on the Blueprint Medicare data can be determined through a comparison between the various levels of risk adjustment and the crude value. Exhibit 12 looks at levels 1 (age +sex) and 3 (HCC scores) of risk adjustment compared to the crude results for both the Blueprint and non-Blueprint Medicare populations. In general, patterns that existed in the Blueprint population also existed in the non-Blueprint population. Any exceptions are described below.

In most cases, level 3 risk adjusters had a greater impact than level 1 adjusters when compared to the crude data. Exceptions to this were in the Blueprint and non-Blueprint results for medical admission, medical inpatient days, outpatient visits to specialists, allowed charges for durable medical equipment (DME), allowed charges for hospice, and allowed charges for long term care. In these cases, the level 1 adjusters had a greater effect on the results. In the case of ASC admissions, level 1 adjusters made a larger difference in the Blueprint results, but did not have the same effect on the non-Blueprint data. For total admissions, level 1 adjusters had a greater impact when compared to level 3 for the non-Blueprint participants, but not for those attributed to Blueprint.

Both level 1 and level 3 adjusters had the same impact on ED allowed charges and ED CMS payments. Although the magnitude was the same for these measures, the effect it had on cost and utilization moved the results in opposite directions. For example, level 1 adjusters in the non-Blueprint population had a percentage difference of -0.4 percent while the level 3 adjusters had a 0.4 percent impact. Conversely, for the Blueprint data, the level 1 adjuster created a 0.5 percent difference in the data and a -0.5 percent difference between level 3 adjusters and the crude.

The largest difference in magnitude occurred for medical admissions, where the difference between the level 1 adjusted results and the crude was -24 percent for non-Blueprint and -25 percent for Blueprint.

Exhibit 12. Effects of Risk Adjustment on Blueprint Medicare Results

		Blueprint		Non-Blueprint	
		Level 1 vs. Crude	Level 3 vs. Crude	Level 1 vs. Crude	Level 3 vs. Crude
Admissions per 1,000: ASC	Abs diff	0.3	2.0	0.11	0.01
	% diff	*-0.50%*	*-3.30%*	*0.20%*	*0.00%*
Admissions per 1,000: Medical	Abs diff	28	3.4	27	5.3
	% diff	*-24.00%*	*2.90%*	*-25.30%*	*5.00%*
Admissions per 1,000: Surgical	Abs diff	0.4	2.5	*<0.1*	1.5
	% diff	*-0.50%*	*-3.10%*	*<0.01%*	*-2.00%*
Admissions per 1,000: Total	Abs diff	12.6	11.9	10.8	15.9
	% diff	*-5.00%*	*4.70%*	*-4.70%*	*6.90%*
IP Days per 1,000: Total	Abs diff	13.5	52.7	12.4	50.8
	% diff	*-0.90%*	*-3.70%*	*-1.00%*	*-4.10%*
IP Days per 1,000: ASC	Abs diff	7.7	18.6	8.7	19.6
	% diff	*-2.50%*	*6.10%*	*-3.60%*	*8.20%*
IP Days per 1,000: Medical	Abs diff	50.4	17	47.4	22
	% diff	*-8.10%*	*2.70%*	*-8.70%*	*4.00%*
IP Days per 1,000: Surgical	Abs diff	39.1	51.6	42	57.1
	% diff	*-7.80%*	*-10.30%*	*-9.40%*	*-12.80%*
Avoidable ED Visits per 100: Without Admission for All Diagnoses	Abs diff	0.05	0.2	0.1	0.3
	% diff	*-0.40%*	*1.20%*	*-0.50%*	*3.10%*
Avoidable ED Visits per 100: Without Admission for Primary Diagnosis	Abs diff	0.1	*<0.1*	0.1	0.2
	% diff	*0.50%*	*0.40%*	*0.80%*	*2.50%*

		Blueprint		Non-Blueprint	
		Level 1 vs. Crude	Level 3 vs. Crude	Level 1 vs. Crude	Level 3 vs. Crude
Number of ED Days per 100: Including Days due to Inpatient Admission	Abs diff	1	2.2	1.6	3.2
	% diff	*1.40%*	*3.00%*	*2.70%*	*5.30%*
Number of ED Days per 100	Abs diff	0.9	1.2	1.5	2
	% diff	*1.60%*	*2.00%*	*3.00%*	*4.10%*
ED: Allowed Charges per Day (Excluding Admissions)	Abs diff	148.8	158.3	148	157.4
	% diff	*-0.40%*	*0.40%*	*0.50%*	*-0.50%*
ED: CMS Payment per Day (Excluding Admissions)	Abs diff	108.8	110.6	108	110
	% diff	*-0.40%*	*0.40%*	*0.50%*	*-0.50%*
Outpatient Visits per 100: Primary Care Providers	Abs diff	9	12.4	9.2	8.9
	% diff	*1.60%*	*2.10%*	*1.40%*	*1.40%*
Outpatient Visits per 100: Specialists	Abs diff	9.4	3.4	8.6	2.1
	% diff	*-2.30%*	*0.80%*	*-2.30%*	*0.60%*
Outpatient Visits per 100: Total	Abs diff	0.8	2.6	1.2	1.7
	% diff	*-0.10%*	*0.30%*	*0.10%*	*-0.20%*
Allowed Charges: Acute Care Hospital Services	Abs diff	7.8	5.6	7.7	5.5
	% diff	*-0.30%*	*0.20%*	*0.30%*	*-0.20%*
Allowed Charges: DME	Abs diff	2	0.9	2	0.9
	% diff	*-0.20%*	*0.10%*	*0.30%*	*-0.10%*
Allowed Charges: E & M	Abs diff	1.6	5.5	1.6	5.5
	% diff	*-0.20%*	*0.50%*	*0.20%*	*-0.50%*
Allowed Charges: Home Health Agency	Abs diff	*<0.1*	4.7	*<0.1*	4.7
	% diff	*<0.01%*	*1.00%*	*0.00%*	*-0.90%*
Allowed Charges: Hospice	Abs diff	0.5	0.4	0.5	0.4
	% diff	*-0.50%*	*0.40%*	*0.70%*	*-0.50%*
Allowed Charges: Imaging	Abs diff	0.2	1.7	0.2	1.7
	% diff	*<0.01%*	*0.30%*	*<0.01%*	*-0.30%*
Allowed Charges: Long Term Care Hospitals	Abs diff	2.4	1.6	2.4	1.6
	% diff	*-0.80%*	*-0.50%*	*1.50%*	*1.00%*
Allowed Charges: Procedures	Abs diff	0.8	2.7	0.8	2.7
	% diff	*-0.10%*	*0.30%*	*0.10%*	*-0.30%*
Allowed Charges: Skilled Nursing Facilities	Abs diff	0.2	7.3	0.2	7.2
	% diff	*0.02%*	*1.18%*	*<0.01%*	*-1.80%*
Allowed Charges: Tests	Abs diff	0.6	0.8	0.6	0.8
	% diff	*-0.20%*	*0.30%*	*0.20%*	*-0.30%*
Allowed Charges: All Services	Abs diff	15.3	29.7	15.3	29.5
	% diff	*-0.20%*	*0.40%*	*0.20%*	*-0.40%*

Blueprint Baseline Quality of Care Assessment by Payer

Medicare Population

Overall, diabetes test results were better for the Medicare Blueprint population than for the non-Blueprint population at baseline: 64.4 percent of all Medicare beneficiaries attributed to the Blueprint population received all three diabetes tests, compared to 60.2 percent in their non-Blueprint counterparts (Exhibit 13).

Multivariate logistic regressions adjusting for age, sex, and year were also run for the Medicare population. Medicare Blueprint participants were more likely to receive A1c testing (odds ratio = 1.63, $p<0.05$) and all three tests (odds ratio = 1.30, $p<0.05$). Women were less likely to receive A1c tests (odds ratio = 0.65, $p<0.05$, results not shown).

Exhibit 13. Blueprint vs. Non-Blueprint Medicare Diabetes Testing Results

	Blueprint[1]	Non-Blueprint[1]	Odds Ratio [2]	Relative risk[2]
HbA1c	93.1%	89. 8%	1.6* (1.02-2.61)	1.0 (.53 – 1.07)
Lipid	85.2%	80.9%	1.3 (0.92-1.84)	1.1 (.99 – 1.1)
Eye exam	78.1%	76.8%	1.3 (0.97-1.75)	1.1 (.99 – 1.1)
Composite	**64.4%**	**60.2%**	**1.3*** (1.01-1.69)	**1.1** (1.0 – 1.19)

1. crude unadjusted rates
2. adjusted for age, sex and year; the non-Blueprint comparison group was the reference group
** statistically significant at p≤0.05*

Commercial Population

For the commercial Blueprint population in Burlington (Exhibit 14), 33.3 percent of participants received all three diabetes exams, which was 2.6 percent higher than the non-Blueprint comparison group. Results were similar between the Blueprint and non-Blueprint groups in St. Johnsbury with 37.5 percent of Blueprint receiving all three tests and 38.1 percent for non-Blueprint. While differences were not statistically significant at the p≤0.05 level, relatively small sample sizes led to wide confidence intervals.

Exhibit 14: Blueprint vs. Non-Blueprint Commercial Diabetes Testing Results

	Burlington		St. Johnsbury	
	Blueprint	**Non-Blueprint**	**Blueprint**	**Non-Blueprint**
Lipid test	74.6% (63.1% - 86.1%)	60.8% (53.1% - 68.6%)	70.8% (59.6% - 82.0%)	77.7% (71.9% - 83.5%)
Eye Exam	39.7% (26.8% - 52.6%)	50.6% (42.7% - 58.5%)	45.8% (33.6% - 58.0%)	47.0% (40.1% - 53.9%)
HbA1c test	82.5% (72.4% - 92.7%)	71.1% (63.9% - 78.3%)	81.9% (72.4% - 91.5%)	86.0% (81.2% - 90.9%)
Composite	**33.3%** (20.9% - 45.8%)	**30.7%** (23.4% - 38.0%)	**37.5%** (25.6% - 49.4%)	**38.1%** (31.4% - 44.9%)

New York City Primary Care Information Project

PCIP and Non-PCIP Population Overview by Payer

Medicare Population

As shown in Exhibit 15, the Medicare population in New York City was much more diverse than Vermont. The patients participating in the PCIP program were statistically significantly different from those in the geographically weighted comparison group on every dimension except gender and COPD diagnosis, however, the magnitude of these differences was small, making the groups similar overall.

The PCIP Medicare population was slightly younger and more racially diverse than the non-PCIP populations based on claims records, with 20.9 percent of participants being recorded as black and 13.9 percent Hispanic compared to 17.9 percent of the non-PCIP population being recorded black and 7.6 percent Hispanic (p<.0001). The PCIP population was also poorer (p<.0001) and lived in zip codes in which residents were slightly less wealthy and less educated compared to the residences of the non-PCIP population (p<.0001 for all levels of education). The PCIP Medicare population had 7.6 percent more participants falling in the lowest quintile for median household income and 4.2 percent fewer participants in the highest quintile according to zip-level median household income from the 2000 U.S. Census (p<.0001).

The PCIP Medicare population generally had lower HCC scores than the non-PCIP population (66.4% of the PCIP population and 63.6% of the non-PCIP population had an HCC score of less than 1; 6.2% of the non-PCIP population had an HCC score above 2.50 (p<.0001)). In the PCIP population, Medicare beneficiaries had a 1.1 percent lower mortality rate (p<.0001) and a 1.5 percent lower rate of congestive heart failure (p<.0001) than the non-PCIP group; however it should also be noted that the PCIP population had 2.3 percent more diabetics (p<.0001) and 2.8 percent more of the population was originally entitled to Medicare for disabilities (p<.0001).

Exhibit 15: PCIP Medicare Population Demographics

	PCIP		non-PCIP		prob.
Person - years (2005-2007)	20,593.3		289,085.4		
Borough[1]	Unwt	Wt	Unwt	Wt	
Bronx	14.2%	14.2%	13.2%	14.2%	
Brooklyn	39.6%	39.6%	29.0%	39.6%	
Manhattan	21.6%	21.6%	24.5%	21.6%	
Queens	18.1%	18.1%	28.0%	18.1%	
Staten Island	6.5%	6.5%	5.3%	6.5%	
PCP assigned[2]		89.1%		79.3%	<.0001
Age					
20-49		7.0%		5.8%	
50-64		10.7%		9.4%	
65-69		20.7%		19.4%	
70-74		18.8%		18.2%	
75-79		16.9%		18.1%	

	PCIP		non-PCIP		prob.
80-99		25.8%		29.2%	<.0001
Female		62.8%		62.2%	0.18
Race recorded in claims records					
Black		20.9%		17.9%	
Hispanic		13.9%		7.6%	
Other or Missing/Unknown[3]		65.2%		74.5%	<.0001
Median income in residential zip code[4]					
lowest quintile		28.4%		20.8%	
2		29.2%		31.4%	
3		19.1%		18.3%	
4		15.8%		17.9%	
highest quintile		7.4%		11.6%	<.0001
Education: Average percent of adults living in patients' residential zip codes with:[5]					
Less than HS		19.0%		17.2%	<.0001
High school diploma		60.0%		58.5%	<.0001
College degree		21.0%		24.4%	<.0001
Mortality (one year)		3.0%		4.1%	<.0001
Diabetes		24.5%		22.2%	<.0001
CHF		8.8%		10.3%	<.0001
COPD		7.3%		8.0%	0.0094
Disabled[6]		26.6%		23.8%	<.0001
HCC score					
< 0.75		50.1%		47.9%	
0.75 - < 1.00		16.3%		15.7%	
1.00 - < 1.50		18.2%		18.3%	
1.50 - < 2.50		10.6%		12.0%	
2.50 +		4.8%		6.2%	<.0001

All variables are weighted by county weight except "Pers-years" (unweighted) and "Borough" (as indicated below).
1. unwt = pop distribution w/o county weighting; wt = pop distrib w. county weighting
2. fraction assigned to primary care physicians (vs. medical / surgical specialists)
3. includes missing / unknown race and white.
4. median household income of bene zipcode based on Census 2000
5. zip-level education statistics based on Census 2000
6. original reason for Medicare entitlement- virtually everyone in Medicare <65 is disabled

Medicaid Population

As was true of the Medicare population, PCIP Medicaid patients differed from their non-PCIP counterparts. These differences were all statistically significant for every characteristic (p<.0001), as

shown in Exhibit 16. However, the magnitude of the differences was small. Thus, by and large, the populations were similar in the PCIP and non-PCIP groups.

The PCIP Medicaid population was slightly older (1% more were between the ages of 35–64) and 1.7 percent fewer were female relative to the comparison group. Educational attainment (as measured by average percent of adults living in patients' residential zip codes with various degrees), income (as measured by median income in patients' residential zip codes) and race (as recorded in claims records) also varied between the PCIP and comparison group. The PCIP group was less wealthy with 1.4 percent more participants in the lowest income quintile and 0.9 percent fewer participants in the highest quintile. For PCIP, 0.5 percent fewer participants lived in a zip code where residents had received either a high school diploma or a college degree. There were also 4.3 percent fewer recorded blacks and 1 percent more Hispanics among the PCIP participants compared to the comparison group. However, since a large proportion of eligibility records were missing any indication of race, it is important to note that the actual racial composition of PCIP and non-PCIP population groups may be different from what is reflected in this analysis.

Health characteristics of the two populations demonstrate that PCIP Medicaid beneficiaries were slightly healthier than the comparison group. Overall HCC scores were slightly lower for the PCIP population with 0.8 percent more participants having an HCC score below 0.75 and 0.6 percent less having an HCC score above 1.0. Additionally, a smaller proportion of PCIP participants had CHF and COPD (0.2% less and 0.4% less respectively), although 0.5 percent more of the PCIP participants had diabetes. Finally, a slightly higher proportion of PCIP Medicaid participants (0.5%) were enrolled in managed care.

Exhibit 16: PCIP and Non-PCIP Medicaid Population Demographics

	PCIP		non-PCIP		prob.
Person - years (2007)	130,313		475,210		
Borough (unwt % / wt %)[1]					
Bronx	13.8	13.8	21.6	13.8	
Brooklyn	41.3	41.3	35.7	41.3	
Manhattan	13.3	13.3	14.7	13.3	
Queens	31.6	31.6	28.0	31.6	
PCP assigned[2] (%)		98.0		95.3	<.0001
Age (%)					<.0001
20-34		37.6		38.6	
35-49		38.7		37.5	
50-64		23.7		23.9	
Female (%)		65.4		67.1	<.0001
Race recorded in claims records[3] (%)					<.0001
Black		21.5		25.8	
Hispanic		9.6		8.6	
Other or Missing/Unkown[4]		68.9		65.6	
Median income in residential zip code[5] (%)					<.0001
lowest quintile		12.7		11.3	
2		44.2		41.8	
3		28.9		29.2	

	PCIP		non-PCIP		prob.
4		11.8		14.6	
highest quintile		2.3		3.2	
Education: Average percent of adults living in patients' residential zip codes with:[6] (%)					
Less than HS		36.8		35.7	<.0001
High school diploma		40.1		40.6	<.0001
College degree		23.2		23.8	<.0001
Diabetes (%)		13.0		12.5	<.0001
CHF (%)		0.9		1.1	<.0001
COPD (%)		6.3		6.7	<.0001
Managed Care (%)		61.0		60.5	<.0001
HCC score (%)					<.0001
< 0.75		90.5		89.7	
0.75 - < 1.00		4.3		4.4	
1.00 - < 1.50		3.5		3.7	
1.50 - < 2.50		1.4		1.6	
2.50 +		0.4		0.6	

All variables are weighted by county weight except "Person-years" (unweighted) and "Borough" (as indicated below).
1. unwt = pop distrib w/o county weighting; wt = pop distrib w. county weighting. Staten Island resdents excluded, since only 2% of PCIP patients lived in the borough.
2. fraction assigned to primary care physicians (vs. medical / surgical specialists)
3. Note that a large proportion of eligibility records were missing indication of race
4. includes missing / unknown race
5. median household income of bene zipcode based on Census 2000
6. zip-level education statistics based on Census 2000
Above samples exclude Medicaid recipients who have been considered managed care exempt (HIV/AIDS, Serious Mental Illness (SMI/SPMI), or OMRDD) as well as excluding SSI recipients and those who are eligible for both Medicaid and Medicare

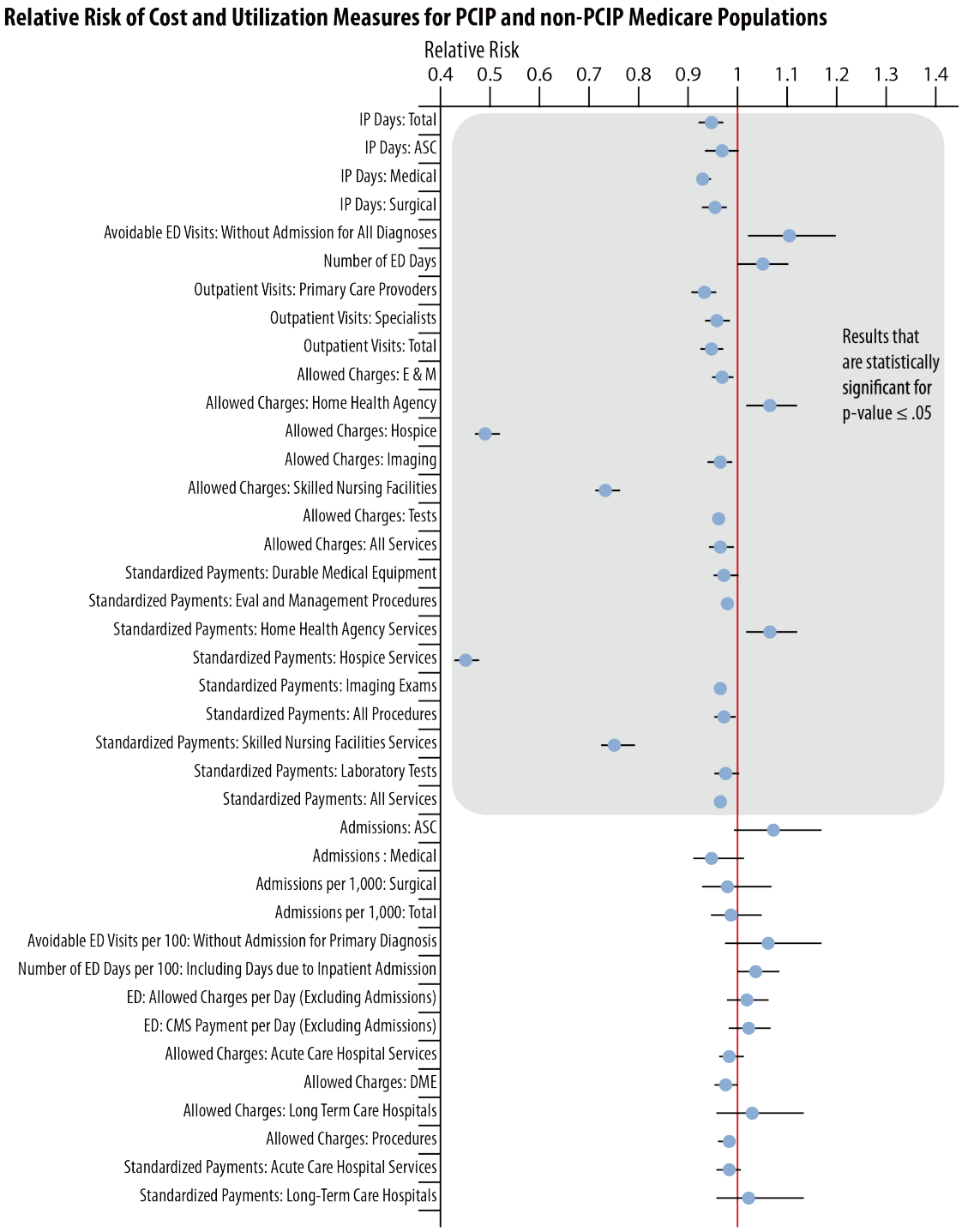
PCIP Baseline Cost and Utilization Assessment by Payer

Medicare Population

Full results can be found in Appendix B3 and Exhibit 17 summarizes the cost and utilization data for the PCIP Medicare population relative to the non-PCIP population after risk adjustment.[34] In only four out of 25 statistically significant results ($p \leq .05$), patients attributed to the PCIP group had higher cost and utilization. For avoidable ED visits and ED days, PCIP had higher utilization rates (9% higher and 4% higher, respectively). For home health agency (HHAs) costs, PCIP also had higher rates: 5 percent higher allowed charges and standardized payments for HHAs. However, PCIP Medicare patients had notably lower costs for hospice care and skilled nursing facilities compared to their non-PCIP counterparts.

[34] Data has undergone level 3 risk-adjustment for PCIP: age, sex, race, education, income and HCC scores.

Exhibit 17. Relative Risk of Cost and Utilization Measures for PCIP vs. non-PCIP Medicare Populations



The relative risk table displays the ratio of something occurring in the intervention group relative to the comparison (non-PCIP) group and therefore demonstrates differences in the baseline data between the two groups. For measures where the relative risk is greater than 1, the event happens more frequently or costs more in the PCIP group. When the relative risk is lower than 1, the opposite is true and the event occurs or costs more in the non-PCIP group. For example, a relative risk of 1.05 means that an event happens or costs 5% more in the PCIP group.

When the confidence interval of a relative risk crosses 1, there were no statistically significant differences between the results for the intervention and comparison groups. The relative risks displayed above have been adjusted using tier 3 methods (ASR + zip income + prior year HCC scores). The bars on each relative risk represent the 95% confidence interval.

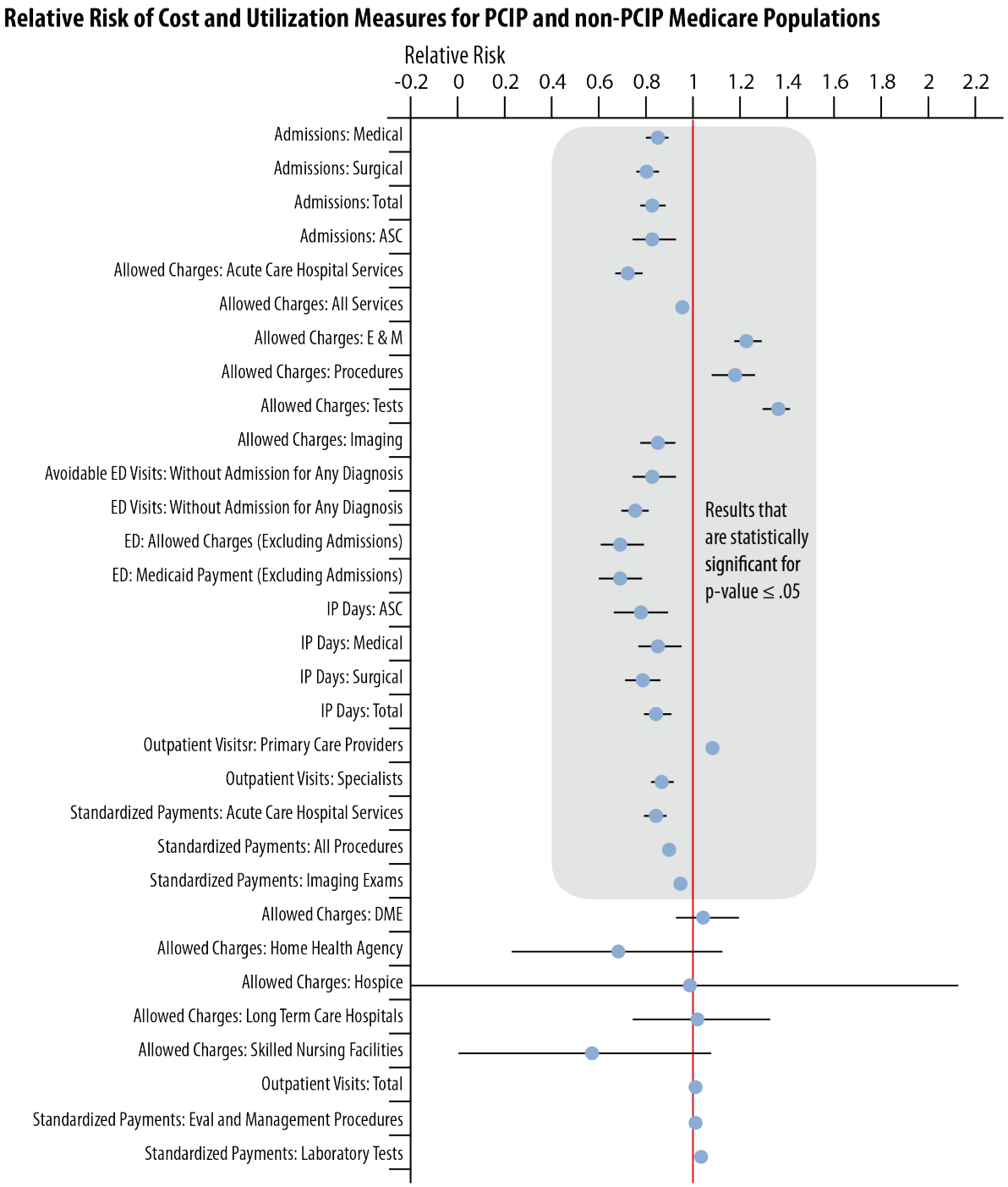
Medicaid Populations

Exhibit 18 summarizes the cost and utilization data for the PCIP Medicaid population relative to the non-PCIP population after risk adjustment. Full data can be found in Appendix B8. In only four out of 23 statistically significant measures ($p \leq .05$), PCIP Medicaid patients had higher rates of cost and utilization.

For physician services, allowed charges for evaluation and management procedures (E&M) and tests were 16 to 36 percent higher for PCIP Medicaid patients compared to their non-PCIP counterparts. However, allowed charges were 27 percent lower for acute care hospital services and 18 percent lower for imaging for the PCIP group. Allowed charges for all services were only 7 percent lower for PCIP participants. While expenditures for allowed charges varied, standardized payments were lower for the PCIP group than the non-PCIP group by 4 to 17 percent for hospital services, imaging exams, and overall all procedures.

PCIP Medicaid patients had 4 percent more outpatient visits to primary care providers and 17 to 21 percent lower rates of admission for medical, surgical and total procedures. ED cost and utilization were 20 to 30 percent lower for the PCIP group. Inpatient days were also lower for PCIP by 11 to 19 percent . Although outpatient visits were higher to primary care providers in the PCIP group, they were 13 percent lower for specialists.

Exhibit 18. Relative Risk of Cost and Utilization Measures for PCIP vs. non-PCIP Medicaid Populations



The relative risk table displays the ratio of something occurring in the intervention group relative to the comparison (non-PCIP) group and therefore demonstrates differences in the baseline data between the two groups. For measures where the relative risk is greater than 1, the event happens more frequently or costs more in the PCIP group. When the relative risk is lower than 1, the opposite is true and the event occurs or costs more in the non-PCIP group. For example, a relative risk of 1.05 means that an event happens or costs 5% more in the PCIP group.

When the confidence interval of a relative risk crosses 1, there were no statistically significant differences between the results for the intervention and comparison groups. The relative risks displayed above have been adjusted using tier 3 methods (ASR + zip income + prior year HCC scores + Medicaid managed care receipt). The bars on each relative risk represent the 95% confidence interval.

Impact of Risk Adjustment by Payer

As noted earlier, the above analysis on baseline costs and utilization was conducted using the highest tier of risk adjustment in each payer population. Separate analysis was undertaken in order to assess the relative impact of different tiers of risk adjustment on statistical significance, magnitude, and direction of relative risk.

Medicare Population

Out of 39 measures, risk adjustment resulted in differences in 12 measures (Exhibit 19). For example, in the case of acute care charges, DME charges and standardized payments for acute care hospital services, results were no longer statistically significant in the third tier of risk adjustment. In the case of total admits, ED allowed charges, ED CMS payments, allowed charges for home health and standardized payments for home health agency services, results were only statistically significant in a second or third tier risk adjustment level.

While risk adjustment methods did not significantly alter the magnitude of measures, in certain cases it did change the directionality of risk although not always to a statistically significant degree. For example, non-PCIP patients were more likely to have higher ED allowed charges, ED CMS payments per day, and home health agency allowed charges when level 1 or level 2 risk adjustment methods were used. However, when level 3 risk adjustment methods were used, non-PCIP patients became less likely to have those higher charges.

Exhibit 19: Statistical Significance of Risk Adjustment Methods on PCIP vs. Non-PCIP Results for Medicare

		Non-PCIP vs PCIP		
		Relative Risk		
		Level 1a/	Level 2 b/	Level 3 c/
Admissions: Medical	**Relative Risk**	**1.07**	**1.05**	**1.04**
	95% Confidence Interval	1.01 - 1.12	1 - 1.1	0.99 - 1.09
	p-value	0.01	0.06	0.09
Admissions: Total	**Relative Risk**	**1.02**	**1.04**	**1.01**
	95% Confidence Interval	0.99 - 1.06	1 - 1.08	0.97 - 1.04
	p-value	0.25	0.03	0.76
Avoidable ED Visits: Without Admission for Primary Diagnosis	**Relative Risk**	**0.91**	**0.96**	**0.94**
	95% Confidence Interval	0.84 - 1	0.88 - 1.04	0.87 - 1.03
	p-value	0.04	0.32	0.17
Number of ED Days	**Relative Risk**	**0.95**	**0.98**	**0.96**
	95% Confidence Interval	0.92 - 0.99	0.94 - 1.02	0.92 - 1
	p-value	0.01	0.33	0.03
ED: Allowed Charges per Day (Excluding Admissions)	**Relative Risk**	**1.00**	**1.03**	**0.99**
	95% Confidence Interval	0.97 - 1.02	1 - 1.05	0.96 - 1.01
	p-value	0.97	0.04	0.23
ED: CMS Payment per Day (Excluding Admissions)	**Relative Risk**	**1.00**	**1.03**	**0.98**
	95% Confidence Interval	0.97 - 1.02	1 - 1.05	0.96 - 1.01
	p-value	0.97	0.04	0.21
Allowed Charges: Acute Care Hospital Services	**Relative Risk**	**1.05**	**1.07**	**1.00**
	95% Confidence Interval	1.02 - 1.06	1.05 - 1.09	0.98 - 1.02
	p-value	<.01	<.01	0.80
Allowed Charges: DME	**Relative Risk**	**1.08**	**1.09**	**1.02**
	95% Confidence Interval	1.07 - 1.1	1.08 - 1.11	1 - 1.03

		Non-PCIP vs PCIP		
		Relative Risk		
		Level 1a/	Level 2 b/	Level 3 c/
	p-value	<.01	<.01	0.09
Allowed Charges: Home Health Agency	**Relative Risk**	**1.01**	**1.02**	**0.95**
	95% Confidence Interval	0.97 - 1.03	0.99 - 1.05	0.91 - 0.98
	p-value	0.73	0.18	<.01
Allowed Charges: Procedures	**Relative Risk**	**1.04**	**1.03**	**1.01**
	95% Confidence Interval	1.03 - 1.05	1.02 - 1.04	1 - 1.02
	p-value	<.01	<.01	0.20
Standardized Payments: Acute Care Hospital Services	**Relative Risk**	**1.06**	**1.08**	**1.01**
	95% Confidence Interval	1.04 - 1.08	1.06 - 1.1	0.99 - 1.03
	p-value	<.01	<.01	0.32
Standardized Payments: Home Health Agency Services	**Relative Risk**	**1.00**	**1.02**	**0.95**
	95% Confidence Interval	0.97 - 1.03	0.99 - 1.05	0.91 - 0.98
	p-value	0.90	0.15	<.01

[a/] The level 1 adjusters include Age (20-49, 50-64,65-69,70-74,75-79, and 80-99), Gender (male and female) and race (white/other/unknown, black, and hispanic)

[b/] The Level 2 adjusters include Education and Income factors in addition to all Level 1 adjusters. Zip code rates of various levels of schooling (less than a high school diploma, high school diploma, and at least a college degree) are used to control for education. Medicaid status at the person-level as well as income quintile distributions at the zip code level is used to control for income.

[c/] Level 3 adjusters include the CMS Hierarchical Condition Category (HCC) risk scores in addition to the Level 1 and Level 2 adjusters. The relevant performance year is used to generate the HCC score.

Exhibit 20 displays the impact that risk adjustment had on the PCIP Medicare data. Similar to Blueprint, the level 3 risk adjusters made a greater difference in the results than the level 1 adjusters. However, in a number of measures, the converse was true. For ASC admissions, total admissions, total inpatient days, ED allowed charges, and ED CMS payments, level 1 adjusters made a larger percentage difference in the results than level 3 in both the PCIP and non-PCIP groups. In the case of avoidable ED visits without admissions for all diagnosis, the PCIP group level 3 had a greater impact than level 1, however, in the non-PCIP group the level 1 adjusters (age +sex + race) had the greatest impact.

In 21 measures, both levels of risk adjustment had a negative percentage difference on the non-PCIP group and a positive percentage difference in the PCIP group. In only two cases (avoidable ED visits without admission for all diagnosis and for primary diagnosis) was the opposite true and both levels of risk adjustment had a negative percentage impact on the results for PCIP and a positive difference in non-PCIP. Risk adjustment had the greatest impact on allowed charges for hospice in level 3 adjusters (18.26%) for the PCIP group and level three adjusters in inpatient days for ASC conditions in both the PCIP (16.52%) and non-PCIP groups (14.84%).

Exhibit 20: Impact of Risk Adjustment Method in PCIP vs. Non-PCIP Results for Medicare

		Non-PCIP		PCIP	
		Level 1 vs. Crude	Level 3 vs. Crude	Level 1 vs. Crude	Level 3 vs. Crude
Admissions : ASC	Abs diff	12.6	8.5	12.7	8.3
	% diff	-12.30%	8.30%	-11.90%	7.90%
Admissions: Medical	Abs diff	0.8	7.7	1.5	7.7

		Non-PCIP		PCIP	
		Level 1 vs. Crude	Level 3 vs. Crude	Level 1 vs. Crude	Level 3 vs. Crude
	% diff	0.30%	3.30%	-0.70%	3.50%
Admissions: Surgical	Abs diff	1.4	11.2	0.9	5.1
	% diff	-1.50%	-11.80%	1.00%	-5.80%
Admissions: Total	Abs diff	45.7	29.3	47.2	31
	% diff	-10.50%	6.80%	-11.40%	7.50%
IP Days: Total	Abs diff	318.6	94.3	296.1	161.2
	% diff	-9.50%	2.80%	-9.90%	5.40%
IP Days: ASC	Abs diff	3.7	112.2	3.2	120.5
	% diff	-0.50%	14.80%	0.40%	16.50%
IP Days: Medical	Abs diff	74.7	100.6	72	123.1
	% diff	-4.50%	6.00%	-4.90%	8.40%
IP Days: Surgical	Abs diff	99.4	114.8	98.5	106.6
	% diff	-10.80%	-12.40%	-12.30%	-13.30%
Avoidable ED Visits: Without Admission for All Diagnosis	Abs diff	0.3	0.2	0.5	0.7
	% diff	3.00%	1.80%	-4.60%	-6.70%
Avoidable ED Visits: Without Admission for Primary Diagnosis	Abs diff	0.5	0.6	0.2	0.2
	% diff	7.50%	9.00%	-2.20%	-2.80%
Number of ED Including Days due to Inpatient Admission	Abs diff	2.4	4	1.1	2.4
	% diff	4.00%	6.50%	1.70%	3.70%
Number of ED Days	Abs diff	1.6	2.4	0.01	0.4
	% diff	4.50%	6.90%	0.00%	1.00%
ED: Allowed Charges (Excluding Admissions)	Abs diff	382.7	242.1	387.1	244.9
	% diff	1.90%	1.20%	-1.90%	-1.20%
ED: CMS Payment (Excluding Admissions)	Abs diff	282.8	171	286	173
	% diff	1.90%	1.10%	-1.80%	-1.10%
Outpatient Visits: Primary Care Providers	Abs diff	7.7	30.1	4.3	19.6
	% diff	1.00%	3.80%	0.50%	2.00%
Outpatient Visits per 100: Specialists	Abs diff	3.1	6.7	17.2	40.5
	% diff	-0.30%	-0.70%	2.20%	5.10%
Outpatient Visits per 100: Total	Abs diff	11.1	23.8	12.2	24.8
	% diff	-0.70%	-1.40%	0.80%	1.60%
Allowed Charges: Acute Care Hospital Services	Abs diff	35.8	151.8	36.2	153.6
	% diff	-0.60%	-2.60%	0.70%	2.80%
Allowed Charges: DME	Abs diff	10.8	35.9	10.8	35.8
	% diff	-1.30%	-4.40%	1.50%	4.90%
Allowed Charges: E & M	Abs diff	17.6	44.3	17.6	44.2
	% diff	-0.90%	-2.20%	0.90%	2.40%
Allowed Charges: Home Health Agency	Abs diff	15.6	31.3	15.7	31.6
	% diff	-2.80%	-5.60%	3.00%	6.00%
Allowed Charges: Hospice	Abs diff	3.4	7.1	3.4	7.2

		Non-PCIP		PCIP	
		Level 1 vs. Crude	Level 3 vs. Crude	Level 1 vs. Crude	Level 3 vs. Crude
	% diff	-3.60%	-7.50%	8.70%	18.30%
Allowed Charges: Imaging	Abs diff	10.4	21	10.4	20.9
	% diff	-1.10%	-2.30%	1.20%	2.50%
Allowed Charges: Long Term Care Hospitals	Abs diff	5.1	9	5.2	9.1
	% diff	0.80%	-1.40%	-0.80%	1.40%
Allowed Charges: Procedures	Abs diff	27.2	50.3	27.1	50.1
	% diff	-1.90%	-3.40%	2.00%	3.70%
Allowed Charges: Skilled Nursing Facilities	Abs diff	30.2	53.6	30.6	54.2
	% diff	-2.80%	-5..0%	4.30%	7.50%
Allowed Charges: Tests	Abs diff	9	18.7	8.9	18.6
	% diff	-1.50%	-3.10%	1.60%	3.40%
Allowed Charges: All Services	Abs diff	150.9	416.5	152.6	421.2
	% diff	-1.10%	-3.00%	1.20%	3.30%
Standardized Payments: Acute Care Hospital Services	Abs diff	33.8	123.8	34.2	125.2
	% diff	-0.90%	-3.20%	1.00%	3.50%
Standardized Payments: Durable Medical Equipment	Abs diff	6.5	25.2	6.4	25.1
	% diff	-1.00%	-3.90%	1.10%	4.30%
Standardized Payments: Eval and Management Procedures	Abs diff	15.7	41	15.6	40.9
	% diff	-0.90%	-2.30%	0.90%	2.40%
Standardized Payments: Home Health Agency Services	Abs diff	13.1	27.89	13.2	28.2
	% diff	-2.40%	-5.10%	2.50%	5.40%
Standardized Payments: Hospice Services	Abs diff	2.8	5.8	2.8	5.8
	% diff	-3.80%	-7.90%	10.40%	21.40%
Standardized Payments: Imaging Exams	Abs diff	8.5	17.4	8.4	17.3
	% diff	-1.10%	-2.30%	1.20%	2.40%
Standardized Payments: Long-Term Care Hospitals	Abs diff	4.7	6.4	4.7	6.5
	% diff	-1.80%	-3.40%	2.0-%	3.70%
Standardized Payments: All Procedures	Abs diff	24.4	45.4	24.3	45.3
	% diff	-1.80%	-3.40%	2.00%	3.70%
Standardized Payments: Skilled Nursing Facilities Services	Abs diff	21.4	38.4	21.6	38.9
	% diff	-2.90%	-5.20%	4.30%	7.70%
Standardized Payments: Laboratory Tests	Abs diff	9.8	20.4	9.7	20.4
	% diff	-1.50%	-3.20%	1.70%	3.50%
Standardized Payments: All Services	Abs diff	127.6	346.2	129	350.1
	% diff	-1.20%	-3.20%	1.30%	3.50%

Medicaid Population

In the PCIP Medicaid data, risk adjustment resulted in differences in six out of 32 measures (Exhibit 21). In the cases of total outpatient visits, allowed charges in home health, allowed charges for skilled

nursing facilities and standard payments for E&M procedures, results were significant at level 1 and 2 but were no longer statistically significant at level 3 of adjustment. Results were statistically significant for level 1 and level 3 adjusters for standard payments for lab tests but were not significant at the second tier of risk adjustment. For the diabetes eye exam measure, results were not significant when adjusting for age, sex and race (level 1 of adjustment), but were significant at levels 2 and 3.

In the case of standard payments for lab tests, risk adjustment changed the directionality of risk. At level 1 of risk adjustment, PCIP patients were less likley to have higher expenditures for lab tests, however, once HCC scores were added to the adjustment, PCIP patients appeared to have higher expenditure costs.

Exhibit 21: Impact of Risk Adjustment Method in PCIP vs. Non-PCIP Results for Medicaid

		PCIP Relative to Non-PCIP		
		Relative Risk[d/]		
		Level 1	Level 2	Level 3
Outpatient Visits per Person-Year: Total	**Relative Risk** Confidence Interval p-value	**0.99** 0.98 - 1.00 0.0017	**0.99** 0.99 - 1.00 0.0380	**1.00** 1.00 - 1.01 0.3466
Allowed Charges per Person-Year: Home Health Agency	**Relative Risk** Confidence Interval p-value	**0.59** 0.26 - 0.91 0.0128	**0.61** 0.28 - 0.95 0.0230	**0.71** 0.35 - 1.07 0.1173
Allowed Charges per Person-Year: Skilled Nursing Facilities	**Relative Risk** Confidence Interval p-value	**0.49** 0.11 - 0.87 0.0080	**0.48** 0.09 - 0.87 0.0094	**0.60** 0.17 - 1.03 0.0678
Standardized Payments per Person-Year: Eval and Management Procedures	**Relative Risk** Confidence Interval p-value	**0.98** 0.97 - 0.98 <.0001	**0.98** 0.98 - 0.99 <.0001	**1.00** 0.99 - 1.00 0.2822
Standardized Payments per Person-Year: Laboratory Tests	**Relative Risk** Confidence Interval p-value	**0.99** 0.98 - 1.00 0.0027	**1.00** 0.99 - 1.01 0.8064	**1.01** 1.01 - 1.02 0.0003
Proportion of Diabetics Receiving an Eye Exam During the Year	**Relative Risk** Confidence Interval p-value	**1.00** 1.00 - 1.01 0.0544	**1.00** 1.00 - 1.01 0.0184	**1.00** 1.00 - 1.01 0.0086

d/ This is the relative risk of the event being measured under PCIP vs. non-PCIP. For example, a relative risk of 1.05 for medical admissions can be read as a 5 percent increased risk of admission in the PCIP population.

Exhibit 22 displays the absolute and percent differences between the crude results and tier 1 and tier 3 risk adjustment in the PCIP Medicaid populations. The absolute difference for level 1 and level 3 adjusters tended to be the same in both the PCIP and non-PCIP groups. Level 3 of risk adjustment made a greater difference in the results, relative to crude, compared to level 1 risk adjusters. The exception to this rule was in allowed charges for tests, in which case the fuller level of risk adjustment actually resulted in a percentage difference lower than the level 1 adjusters (level 1 adjusters for PCIP had a percentage difference of 0.73% whereas the level 3 adjusters made a difference of 0.73%; results were similar for non-PCIP). In some cases both levels of risk adjustment made a difference that was ≤0.001 percent. This was true in both the PCIP and non-PCIP groups for avoidable ED visits without admission

for diagnosis, avoidable ED visits without admissions for any diagnosis and outpatient visits to specialists. Other instances where risk adjustment made ≤0.0 percent were outpatient visits for primary care and total outpatient visits in the non-PCIP group. Allowed charges for hospice and standard payments for imaging in non-PCIP and allowed charges for DME in PCIP were the only measures for which level 1 risk adjustment made ≤0.0 percent difference and level 3 adjusters did make a difference. However, in the cases of allowed charges for hospice and DME, the absolute difference was still under $1.00 per person-year. In only three cases did risk adjustment make more than a 10% difference from the crude rate (allowed charges for home health, hospice, and skilled nursing facilities). Still, the absolute differences in the results were all under $1 per person-year.

Exhibit 22: Impact of Risk Adjustment Method in PCIP vs. Non-PCIP Results for Medicaid

		Non-PCIP		PCIP	
		Level 1 vs. Crude	Level 3 vs. Crude	Level 1 vs. Crude	Level 3 vs. Crude
Admissions per 1,000 Person-Years[e/]: ASC[g/]	Abs diff	0.2	0.7	0.2	0.7
	% diff	-1.10%	-4.00%	1.60%	5.40%
Admissions per 1,000 Person-Years: Medical	Abs diff	1.4	3.1	1.5	3.2
	% diff	-1.10%	-2.40%	1.50%	3.20%
Admissions per 1,000 Person-Years: Surgical	Abs diff	0.8	2.7	0.9	2.8
	% diff	-0.90%	-3.00%	1.40%	4.30%
Admissions per 1,000 Person-Years: Total	Abs diff	2.3	5.9	2.3	5.9
	% diff	-1.10%	-2.70%	1.40%	3.60%
IP Days per 1,000 Person-Years: ASC[g/]	Abs diff	0.7	3.6	0.7	3.6
	% diff	-1.00%	-5.40%	1.50%	7.60%
IP Days per 1,000 Person-Years: Medical	Abs diff	6.3	20.5	6.2	20.6
	% diff	-1.20%	-3.60%	1.40%	4.60%
IP Days per 1,000 Person-Years: Surgical	Abs diff	4	19.6	4	19.5
	% diff	-1.00%	-5.00%	1.40%	7.00%
IP Days per 1,000 Person-Years: Total	Abs diff	10.2	-40	10.1	40.1
	% diff	-1.10%	-4.30%	1.40%	5.50%
Avoidable ED Visits[h/] per Person-Year: Without Admission for Any Diagnosis	Abs diff	<0.01	<0.01	<0.01	<0.01
	% diff	<0.01%	<0.01%	<0.01%	<0.01%
ED Visits per Person-Year: Without Admission for Any Diagnosis	Abs diff	<0.01	<0.01	<0.01	<0.01
	% diff	<0.01%	<0.01%	<0.01%	<0.01%
ED: Allowed Charges per Person-Year (Excluding Admissions)	Abs diff	0.2	-0.4	0.2	0.3
	% diff	-1.70%	-3.40%	2.60%	4.00%
ED: Medicaid Payment per Person-Year (Excluding Admissions)	Abs diff	0.2	-0.4	0.2	0.3
	% diff	-1.70%	-3.50%	2.70%	4.00%
Outpatient Visits per Person-Year: Primary Care Providers	Abs diff	<0.01	<0.01	<0.01	0.1
	% diff	<0.01%	<0.01%	<0.01%	2.04%
Outpatient Visits per Person-Year: Specialists	Abs diff	<0.01	<0.01	<0.01	<0.01
	% diff	<0.011%	<0.01%	<0.01%	<0.01%
Outpatient Visits per Person-Year: Total	Abs diff	<0.01	<0.01	<0.01	<0.01

		Non-PCIP		PCIP	
		Level 1 vs. Crude	Level 3 vs. Crude	Level 1 vs. Crude	Level 3 vs. Crude
	% diff	<0.01%	<0.01%	<0.01%	1.64%
Allowed Charges per Person-Year: Acute Care Hospital Services	Abs diff	9.9	29.2	9.9	29
	% diff	-1.20%	-3.60%	1.80%	5.30%
Allowed Charges per Person-Year: DME	Abs diff	0.1	0.1	<0.01	0.1
	% diff	-2.20%	-2.20%	<0.01%	2.30%
Allowed Charges per Person-Year: E & M[/]	Abs diff	1.8	2.4	1.8	2.3
	% diff	-2.60%	-3.50%	2.30%	2.90%
Allowed Charges per Person-Year: Home Health Agency	Abs diff	0.1	0.5	0.1	0.5
	% diff	-1.90%	-9.30%	3.30%	16.70%
Allowed Charges per Person-Year: Hospice	Abs diff	<0.01	0.1	<0.01	0.2
	% diff	<0.01%	-10.00%	<0.01%	28.60%
Allowed Charges per Person-Year: Imaging	Abs diff	0.1	0.3	0.1	0.3
	% diff	-1.10%	-3.30%	1.40%	4.30%
Allowed Charges per Person-Year: Long Term Care Hospitals	Abs diff	0.2	0.4	0.1	0.3
	% diff	-1.70%	-3.50%	0.90%	2.80%
Allowed Charges per Person-Year: Procedures	Abs diff	0.6	0.9	0.6	0.9
	% diff	-2.50%	-3.80%	2.30%	3.50%
Allowed Charges per Person-Year: Skilled Nursing Facilities	Abs diff	0.2	0.8	0.2	0.9
	% diff	-2.00%	-8.20%	4.40%	20.00%
Allowed Charges per Person-Year: Tests	Abs diff	0.4	0.3	0.4	0.2
	% diff	-1.90%	-1.50%	1.50%	0.70%
Allowed Charges per Person-Year: All Services	Abs diff	12.8	25.9	12.8	26
	% diff	-0.30%	-0.60%	0.30%	0.60%
Standardized Payments per Person-Year: Acute Care Hospital Services	Abs diff	10.7	49.8	10.6	49.6
	% diff	-0.80%	-3.90%	1.10%	5.10%
Standardized Payments per Person-Year: Eval and Management Procedures	Abs diff	1.8	8.7	1.8	8.7
	% diff	-0.30%	-1.30%	0.30%	1.30%
Standardized Payments per Person-Year: Imaging Exams	Abs diff	<0.01	6.7	<0.01	6.6
	% diff	<0.01%	-1.70%	<0.01%	1.80%
Standardized Payments per Person-Year: All Procedures	Abs diff	0.2	5.4	0.2	5.4
	% diff	0.10%	-1.50%	-0.10%	1.60%
Standardized Payments per Person-Year: Laboratory Tests	Abs diff	0.8	4.5	0.7	4.5
	% diff	-0.20%	-1.30%	0.20%	1.30%

[a/] The level 1 adjusters include age category (20-34, 35-49, and 50-64), gender (male and female) and race/ethnicity (white/other/unknown, black, and Hispanic)

[b/] The Level 2 adjusters include Education and Income factors in addition to all Level 1 adjusters. Zipcode rates of three levels of schooling (less than a high school diploma, high school diploma, and at least a college degree) were used to control for education. Median income quintiles at the zip code level were used to control for income.

[c/] Level 3 adjusters include the CMS Hierarchical Condition Category (HCC) risk score and Medicaid managed care receipt, in addition to the Level 1 and Level 2 adjusters. Year 2007 was used to generate the HCC score.

[d/] This is the relative risk of the event being measured under PCIP vs. non-PCIP. For example, a relative risk of 1.05 for medical admissions can be read as a 5 percent increased risk of admission in the PCIP population.

[e/] Each person had to be enrolled for at least 9 months in Medicaid during the year, and not be covered by SSI, considered "managed care exempt," or be in receipt of Medicare simultaneouly (dual-eligible), to be eligible for the analysis

[f/] Non-PCIP includes Medicaid enrollees who were not attributed to PCIP providers, but who resided in a New York City borough in which at least 5 percent of PCIP patients resided.

Baseline Quality of Care Assessment by Payer

Medicare Population

As shown in Exhibit 23, the diabetic PCIP and non-PCIP Medicare populations had similar results for diabetes testing (50.2% for PCIP and 50.1% for non-PCIP for the composite diabetes measure). However, PCIP participants still had approximately $300 less in overall expenditures over the course of a year. Thus, it would appear that PCIP Medicare patients with a common chronic condition received a similar quality of care for lower costs relative to their non-PCIP counterparts.

Exhibit 23. PCIP vs. Non-PCIP Medicare Diabetes Testing Results

	PCIP[(1)]	Non-PCIP[(1)]	Odds Ratio [(2)]	Relative risk [(2)]
HbA1c	78.77%	76.15%	1.16* (1.05-1.27)	1.03* (1.01 – 1.05)
Lipid	79.72%	78.38%	1.08 (0.98-1.20)	1.02 (1 – 1.04)
Eye exam	66.30%	67.11%	0.98 (0.90-1.06)	0.99* (.97 – 1.02)
Composite	**50.15%**	**50.08%**	**1.02 (0.94-1.10)**	**1.08* (1.06 – 1.1)**

1, crude unadjusted rates
2. adjusted for age, sex and year; the non-PCIP comparison group was the reference group
** statistically significant at p≤0.05*

In multivariate logistic regression analysis adjusting for demographic differences, PCIP patients were more likely to receive an A1c test but were not statistically significantly more likely to get the other tests or all three. Females and blacks were statistically significantly less likely to receive these tests. For example, women were less likely to receive eye exams (odds ratio = 0.81, $p<0.05$) or all three tests (odds ratio = 0.86, $p<0.05$). Blacks were less likely to receive any of the tests (odds ratio = 0.71 for A1c test, 0.68 for eye exams, 0.57 for lipid tests, all $p<0.05$) or all three (odds ratio = 0.63, $p<0.05$). Being older, however, increased the odds of receiving eye exams (odds ratio = 1.26, $p<0.05$), lipid tests (odds ratio = 1.12, $p<0.05$), and all three tests (odds ratio = 1.17, $p<0.05$).

Medicaid Population

As demonstrated in Exhibit 24, PCIP Medicaid diabetes patients had similar results for receipt of an eye exam as their non-PCIP counterparts with 1 percent more patients receiving the exam. The differences in these results were not found to be statistically significant. While PCIP patients had only a slightly higher score for quality of care, the cost of care was significantly lower for PCIP. PCIP patients had lower expenditures for both total allowed charges and standardized payments for all procedures, receiving a similar quality of care for a lower price.

Exhibit 24. PCIP vs. Non-PCIP Medicaid Diabetes Testing Results

	PCIP[(1)]	Non-PCIP[(1)]	Relative risk [(2)]
Eye exam	98%	97%	1.00 (1.00 – 1.01)

1-crude unadjusted rates
2-adjusted for age, sex and year; the non-PCIP comparison group was the reference group

Wisconsin Health Information Organization

Population Overview

As a state, Wisconsin is largely homogenous. Compared nationally, the Wisconsin population is primarily white (86%) and slightly older than the rest of the nation (13.5% of the Wisconsin population is over the age of 65, compared to 12.9% nationally). The median household income in Wisconsin ($49,994 per year) is comparable to the national average, but with fewer people falling under the poverty line than national averages. [35]

Overall the sample for the non-Medicare populations was 67.26 percent commercial subscribers, 24 percent Medicaid beneficiaries and 8.73 percent Medicare Advantage enrollees from 2009 WHIO Data Mart data. All were residents of Wisconsin and must have been enrolled for 9 or more months during the measurement year (except in cases of death). The commercial and Medicaid populations include those under the age of 65 and the sample Medicare Advantage population includes Medicare beneficiaries 20-years-old or over.

As previously discussed, the analysis in Wisconsin was intended to help identify priority areas for future reform efforts rather than evaluate an intervention through an examination of cases and controls. The WHIO analysis allows for the evaluation of the entire enrolled population whereas the Blueprint and PCIP analyses were restricted to beneficiaries with at least one outpatient visit to allow for assignment.

This section begins with a univariate analysis of cost, quality, and utilization by payer, first at the state level and then at the county level. A bivariate, county-level analysis of cost relative to select care quality metrics is then provided to help identify counties that appear to be providing high- or low-value care.

Statewide Payer Averages for Cost, Quality, and Utilization Measures

Total cost of care: For Wisconsin, total cost of care is broken down by payer. The average total cost of care for the Medicare population was $640 per member per month (PMPM). For the commercial and Medicaid populations the total cost was $266 PMPM and $589 PMPM, respectively. Cost was highest for the Medicare Advantage[36] population at $926 PMPM.

ASC admissions: Median ASC admissions were 58.9 admissions per 1,000 person months for the Medicare population. Medicare Advantage had a median ASC admissions rate of 31.5 admissions per 1,000 person months and Medicaid's median rate was 14.5 admissions per 1,000 person months. The commercial population saw the lowest rate of ASC admissions at 2.6 admissions per 1,000 person months.

[35] All data taken from U.S. Census website. http://www.census.gov/
[36] All Medicare Advantage measures include patients 20 and older, which may contribute to higher costs and utilization. Prescription drug costs are also included as part of this measure, which also helps to account for the higher cost.

*HEDIS diabetes score*s: The commercial population performed the best on HEDIS diabetes results with 69 percent of patients receiving all three tests (HbA1c, blood lipid tests and eye exams). The Medicare and Medicare Advantage populations had a rate of 58 percent and 63 percent of patients receiving all three exams, respectively. Only 52 percent of the Medicaid population received all three exams in WHIO.

County-Level Total Cost Across Payers Relative to the Payer mean

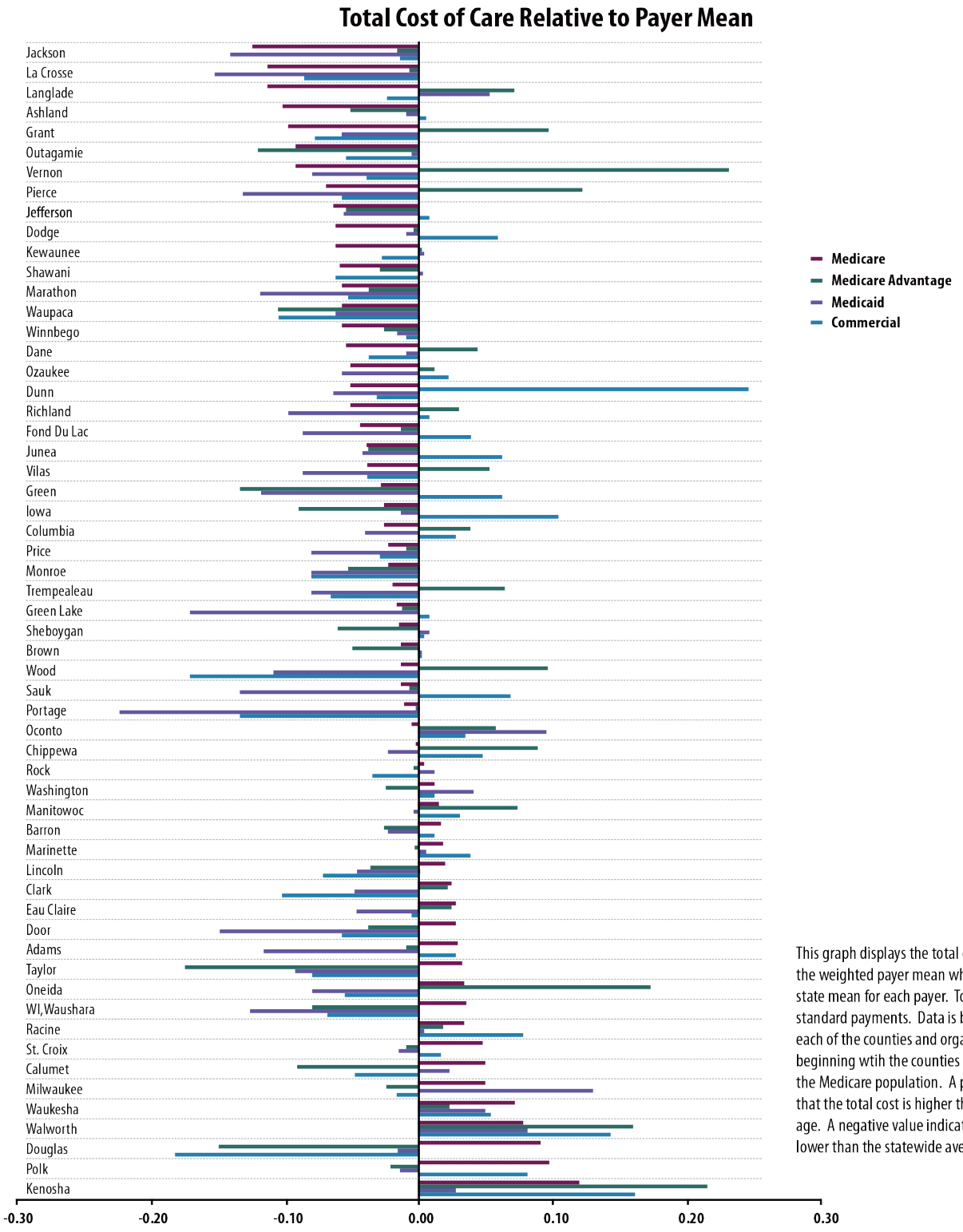
Exhibit 25 displays the risk-adjusted[37] total cost of care[38] for each payer relative to the payer mean, broken down by county. For example, a payer with a score of 1.1 has a total cost that is 10 percent higher than the payer mean for the state. A payer with a score of 0.85 has costs that are 15 percent lower than the state mean for that payer. There are nine counties (Jackson, La Crosse, Outagamie, Marathon, Waupaca, Winnebago, Price, and Monroe) in Wisconsin for which total cost for all four payers are below the payer average for the state. In four counties (Kenosha, Walworth, Waukesha and Racine), total cost for all four payers are above the payer average for the state.

The remaining counties represent areas where at least one payer is above or below the state average while the other payers for that county are not. For example, in Vernon County in Wisconsin, Medicare, Medicaid and the commercial payers are all performing at lower-cost than the payer average. However, the total cost of care for Medicare Advantage in Vernon County is nearly 30 percent higher than the payer mean.

[37] Data are risk adjusted using level 3 adjusters for the Medicare population (age, sex, race, income education and HCC scores) and using an ERG risk adjuster for WHIO

[38] Total cost of care for the WHIO population is defined by total expenditures per person-month; for the non-Medicare data total expenditures include pharmacy data

Exhibit 25. Total Cost Relative to Payer Mean



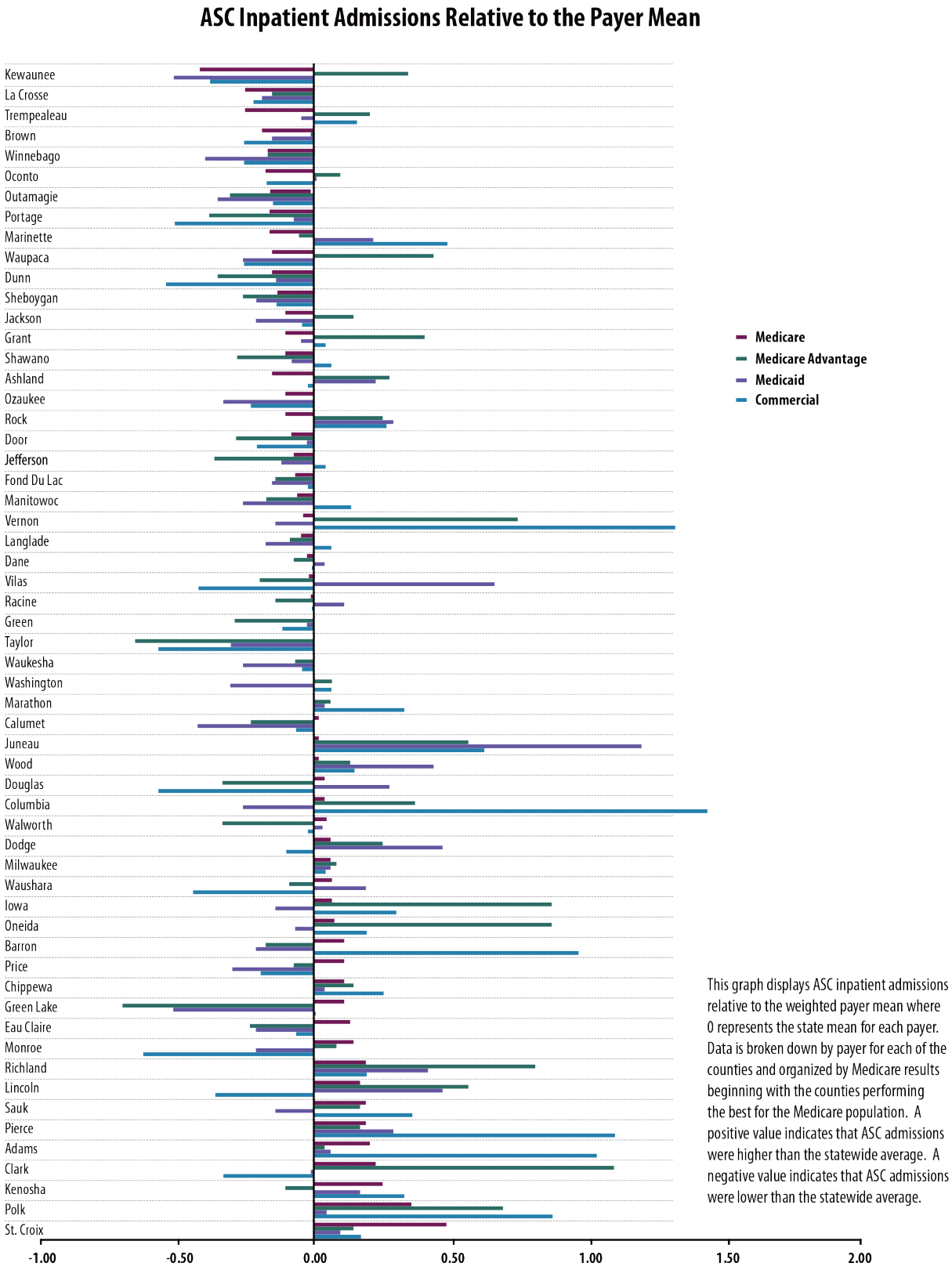
This graph displays the total cost of care relative to the weighted payer mean where 0 represents the state mean for each payer. Total cost is measured in standard payments. Data is broken down by payer for each of the counties and organized by Medicare results beginning wtih the counties performing the best for the Medicare population. A positive value indicates that the total cost is higher than the statewide average. A negative value indicates that the total cost is lower than the statewide average.

County-Level ASC Inpatient Admissions Relative to Payer Mean

Care quality also tends to vary by payer population within counties. Exhibit 26 demonstrates that 13 counties had below-average ASC admission rates for all payers and 10 counties had higher rates of admissions than the state average for all payers.[39] In the vast majority of the counties, ASC admissions were higher than average for some payers and lower than average for others. For example, Kewaunee County outperformed all other counties in ASC admissions for the Medicare population, while Kewaunee's Medicare Advantage population had nearly 50 percent more ASC inpatient admissions than the Wisconsin Medicare Advantage average.

[39] Data are risk adjusted using level 3 adjusters for the Medicare population (age, sex, race, income, education, and HCC scores) and using an ERG risk adjuster for the PCIP.

Exhibit 26. ASC Inpatient Admissions Relative to Payer Mean



This graph displays ASC inpatient admissions relative to the weighted payer mean where 0 represents the state mean for each payer. Data is broken down by payer for each of the counties and organized by Medicare results beginning with the counties performing the best for the Medicare population. A positive value indicates that ASC admissions were higher than the statewide average. A negative value indicates that ASC admissions were lower than the statewide average.

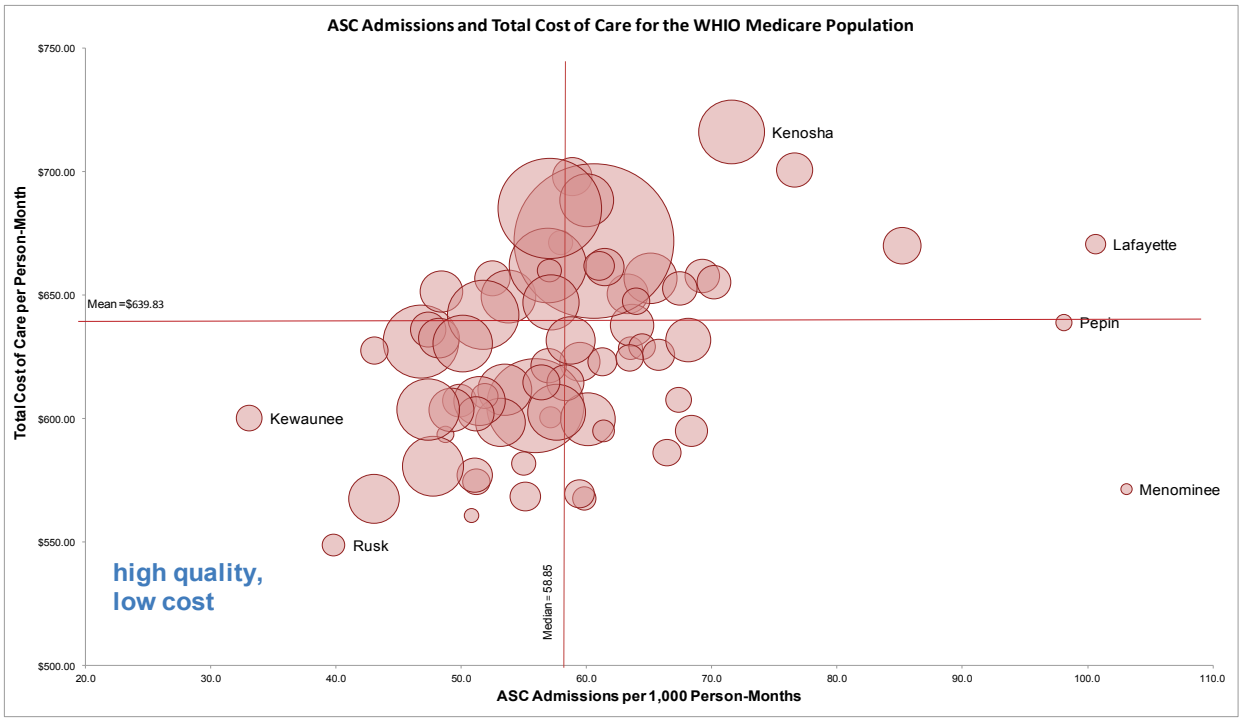
Value of Care in the WHIO Population by Payer

In order to identify those counties that are providing the highest quality health care at the lowest cost, the following analyses chart quality of care (as assessed through ASC admissions and a composite diabetes HEDIS measure) relative to total cost at the county-level by payer. The cross-hairs represent the statewide weighted mean for standard payments and weighted median results for quality. The mean was used for cost measures to take into account total costs while the median was used for quality measures to eliminate the influence of outliers. The data points are marked using circles, with the size of the circles corresponding to the population size in this study for each county.

Medicare

When examining ASC admissions relative to cost, 30 of the 71 counties appear to be providing high-value care as defined by their placement in the "high quality, low cost" quadrant in Exhibit 27. The weighted total cost average for the Medicare population was $640 PMPM and the weighted average of ASC admissions was 58.9 per 1,000 person-months. Kewaunee County had the lowest rate of admission per 1,000 for ASC conditions at 72 per 1,000 person-months (56% of the state Medicare average) with a cost 94 percent of the state average. Kenosha County had the highest total cost at $716 per 1,000 person-months (112% of the statewide mean for Medicare cost) and an admissions rate of 71.6 (122% of the statewide Medicare average). Menominee, Lafayette and Pepin counties exceed the Medicare average than any other counties with ASC admission rates of 103.1, 100.7 and 98.1 per 1,000 person-months, respectively. Menominee's rate of admissions was 175 percent of the statewide mean—Lafayette and Pepin were 171 percent and 167 percent of the WHIO average, respectively.

Exhibit 27. ASC Admissions and Total Cost for WHIO Medicare

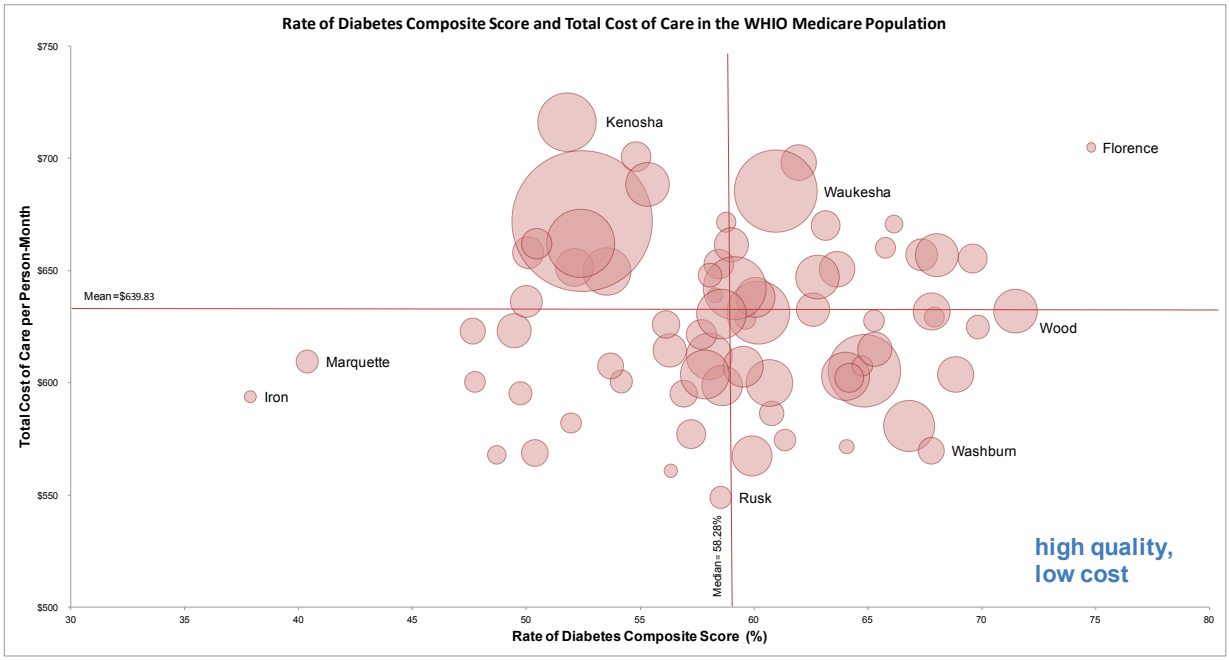


The cost of care for the WHIO Medicare population averaged $640 PMPM with a median composite score of 59 percent for the HEDIS diabetes exams. As displayed in Exhibit 28, 27 counties fell into the

"high quality, low cost" range. Washburn is particularly notable for having the lowest cost in its quadrant ($570 PMPM or 89% of the statewide mean) while also having one of the highest composite scores (68% or 117% of the statewide mean). Outagamie and Menominee reported similar results demonstrating a total cost of $581 PMPM (90% of the WHIO mean) and $571 PMPM (89% of the WHIO mean) with a HEDIS composite score of 67 percent (115% of the WHIO median) and 64% (110% of the median), respectively.

Iron and Marquette counties reported lower than average cost with lower than average composite scores, while Florence reported a composite score 128 percent of the reported state average but also a cost that was 110 percent of the statewide mean.

Exhibit 28. Composite HEDIS Diabetes and Total Cost for WHIO Medicare[40]

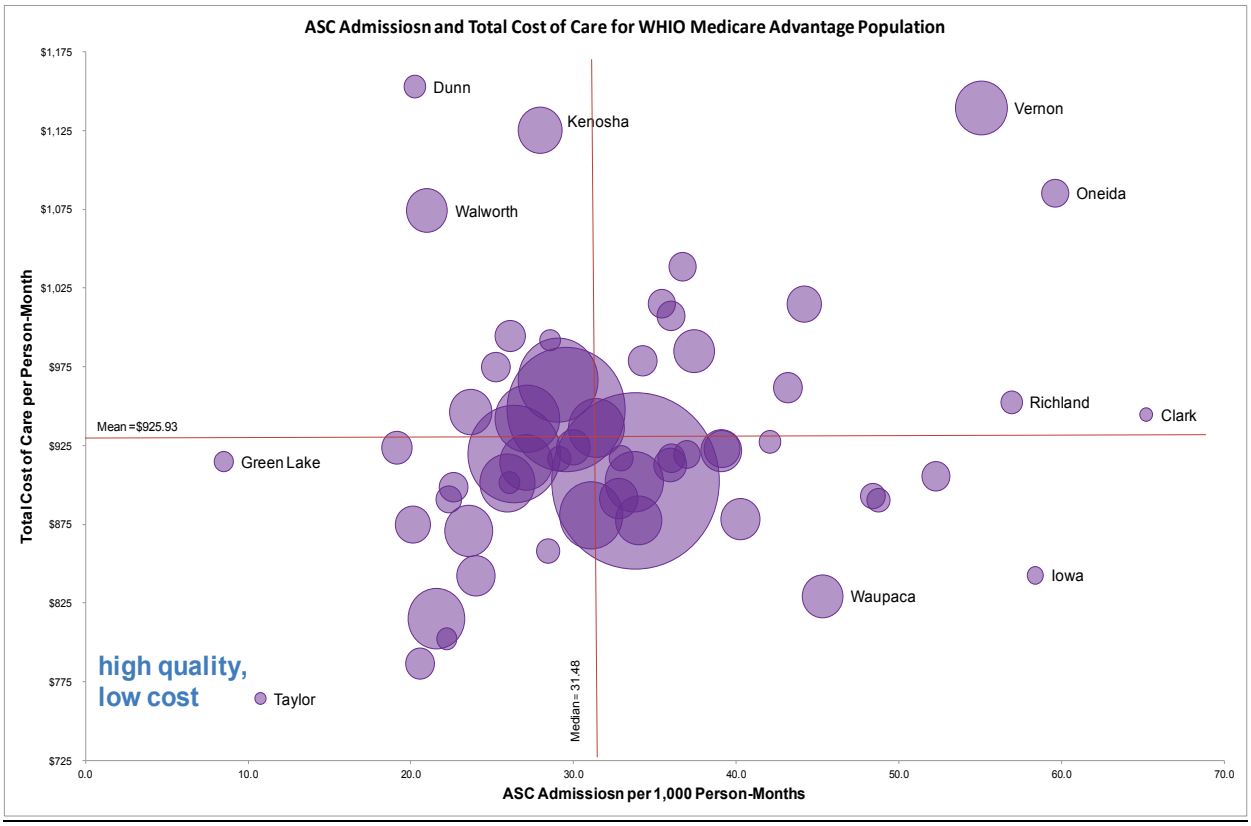


Medicare Advantage

The Medicare Advantage population represents the highest average total cost of care among payers ($926 PMPM) and, second to Medicare, the highest admissions (31.5 admissions per 1000 person months). Nineteen counties fall under the mean cost of care and median number of admissions (Exhibit 29). One noteworthy outlier, Dunn County, while having the highest total cost of care ($1,153 PMPM), had one of the lowest reported ASC admissions (20.2 admissions per 1000 person months) of all counties. Dunn's total cost was 124 percent of Wisconsin's average cost but 64 percent of the statewide ASC admissions. While Clark County's total cost was 102 percent of the Medicare Advantage statewide mean, its ASC admissions rate was 207 percent of the Wisconsin average. Taylor County had the lowest cost (83% of the Wisconsin Medicare Advantage average) and second lowest rate of ASC admissions (only 34% of the average rate).

[40] Counties determined to have a small sample size are those for which N ≤ 100 for the HEDIS diabetes composite measure. Seven counties in the Medicare population were below this threshold (Crawford, Florence, Iron, Jackson, Lafayette, Menominee and Pepin).

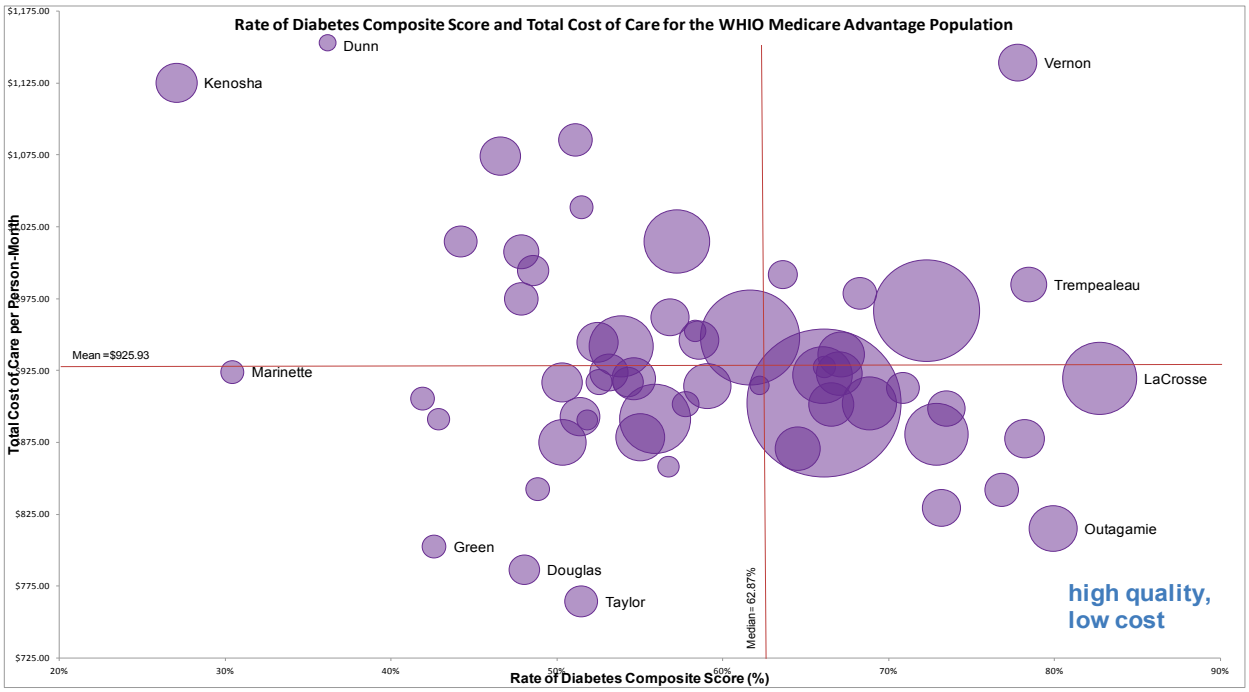
Exhibit 29. ASC Admissions and Total Cost for WHIO Medicare Advantage[41]



The Medicare Advantage population had the highest average total cost of care among payers ($926) and, second to Medicaid, the lowest median composite score, with 63 percent of patients receiving all three of the selected diabetes tests. Within the Medicare Advantage population, variance between counties in cost and quality was particularly high (Exhibit 30). La Crosse County had the highest score in the Medicare Advantage population for the diabetes composite measure with 83 percent of participants receiving all three tests, while Kenosha county had the lowest score at 27 percent (131% and 43% of the statewide Medicare Advantage average). Not only was Kenosha county particularly poor in quality, it also stands out for having the third highest cost in the Medicare Advantage population at $1,125 per person-month (121% of the statewide Medicare Advantage mean). By contrast, Outagamie County provided care at a cost 85 percent of the average while attaining a composite score that was 127 percent of the average.

[41] Counties determined to have a small sample size are those for which N ≤ 200 for the ASC admission measure. Three counties in the Medicare Advantage population were below this threshold (Clark, Iowa and Taylor).

Exhibit 30. HEDIS Composite and Total Cost for WHIO Medicare Advantage [42]

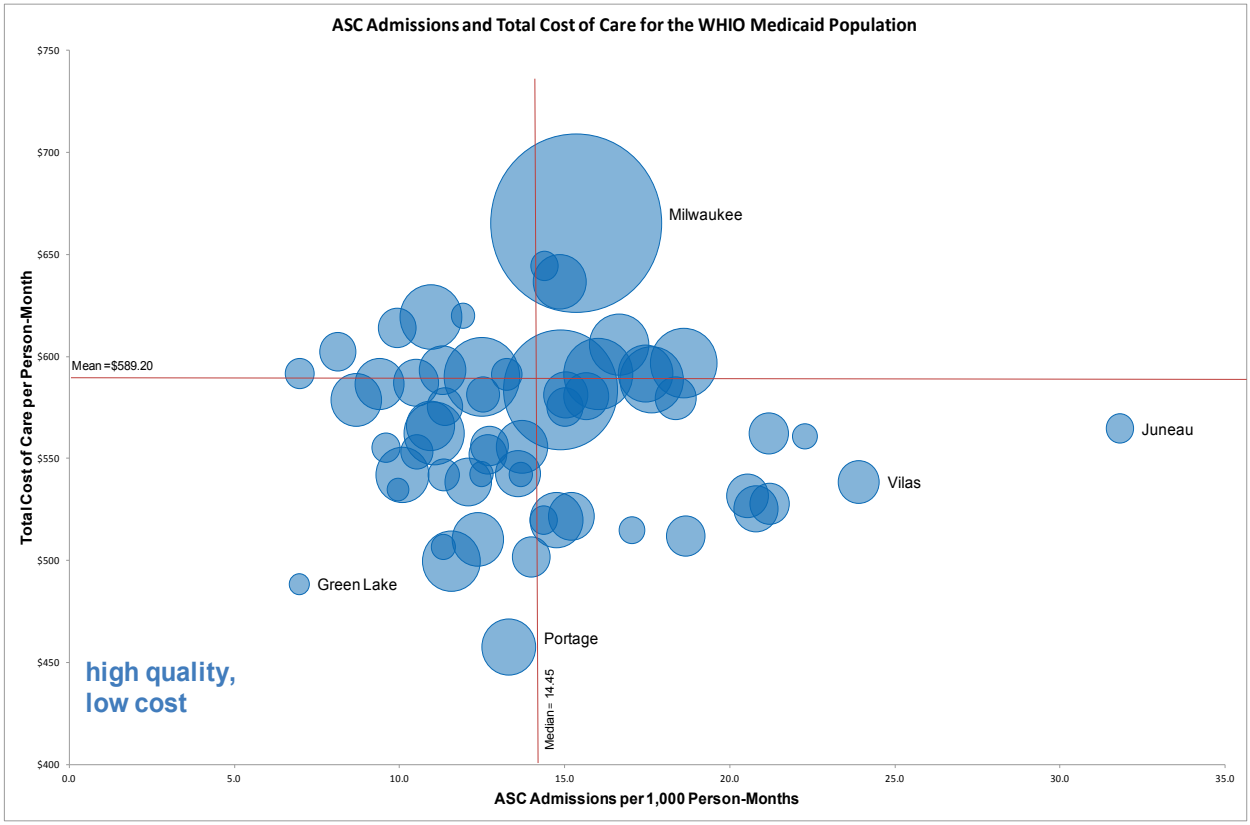


Medicaid

As shown in Exhibit 31, the average cost for the Wisconsin Medicaid population was $589 per person-month and the median number of ASC admission was 14.5 per 1,000 person-months. Several outlier counties are worth highlighting in the WHIO Medicaid population. Green Lake stands out among the high –value counties with the second lowest total cost of care ($488 PMPM; 82% of the WHIO Medicaid average) and tied—with Kewaunee—for the lowest ASC Admission (7.0 admissions per 1,000 person months; only 48% of the state average). The only other county with a lower total cost of care, Portage ($457), had a higher ASC admissions rate, but was still lower than the Medicaid average (91% of the median). Milwaukee had the highest total cost of care for the Medicaid population ($665 PMPM) at 112 percent of the state average and Juneau had the highest number of ASC admissions (31.8 admissions per 1,000 person months) at 220 percent of the statewide Medicaid mean.

[42] Counties determined to have a small sample size are those for which N ≤ 100 for the HEDIS diabetes composite measure. Fourteen counties in the Medicare Advantage population were below this threshold (Barron, Door, Dunn, Green, Green Lake, Iowa, Juneau, Kewaunee, Marinette, Pierce, Polk, Richland, St. Croix and Waushara). Additionally, 10 counties had an N between 100–150.

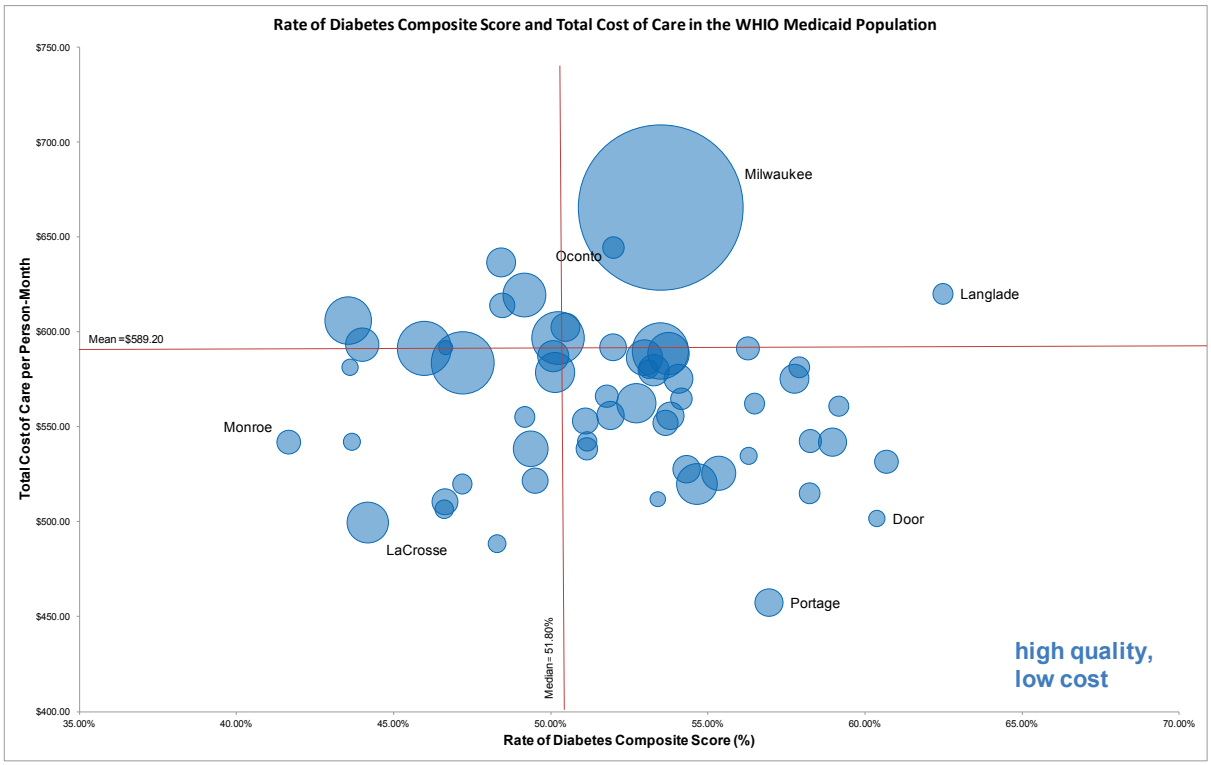
Exhibit 31. ASC Admissions and Total Cost for WHIO Medicaid



The statewide median composite diabetes score for the Medicaid population was 51.8 percent, the lowest among the different paper groups participating in WHIO (Exhibit 32).[43] Portage County again stands out as a high-value county with a total cost of $457 PMPM and a composite diabetes score 110 percent of the statewide Medicaid average. Monroe County had the lowest quality score at 42 percent, which was 81 percent of the Medicaid average.

[43] It is important to note that the small Ns may lead to unstable results.

Exhibit 32. HEDIS Composite and Total Cost for WHIO Medicaid[44]

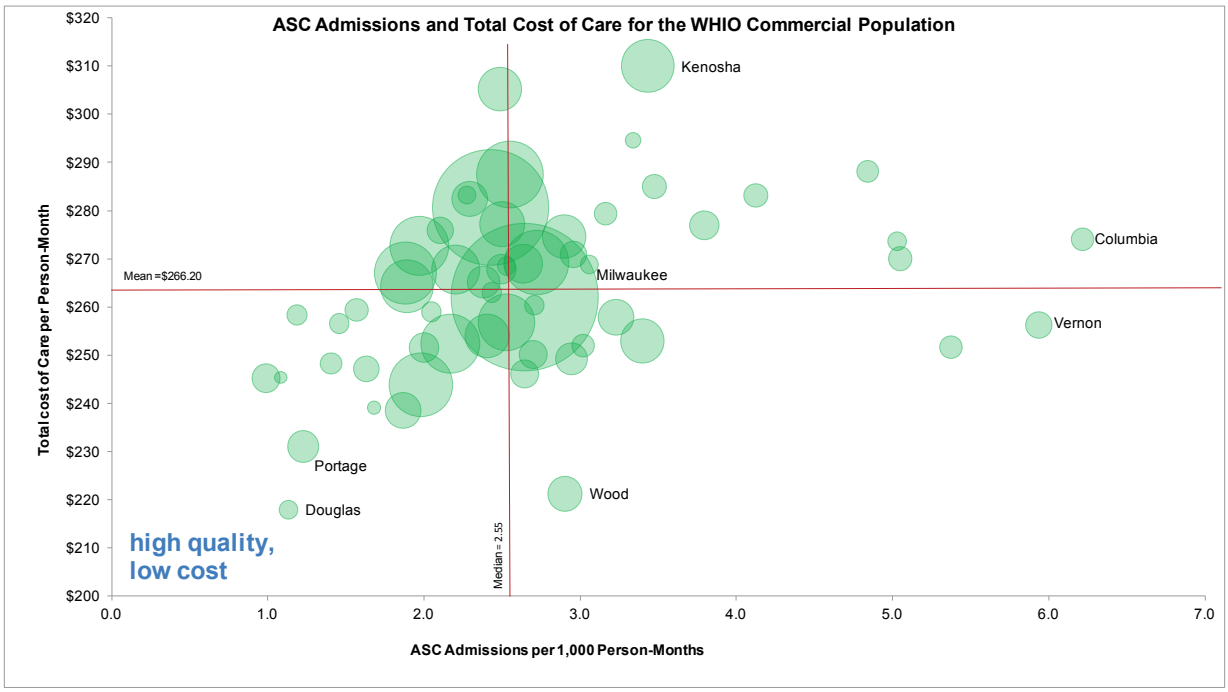


Commercial

The mean cost of care for the commercial population was $266 PMPM, substantially lower than the cost of the other three payers. Similarly, the median ASC Admissions for the commercial population of 2.5 admissions per 1,000 person months was much lower than the other payers. Seventeen counties fell into the "high quality, low cost range," as defined by having total cost and ASC Admissions below the population mean and median (Exhibit 33), of which Douglas (82% of the average cost, 43% of the average ASC admissions) and Portage (87% of the average cost, 47% of the average admissions) counties stand out. Conversely, Kenosha County had the highest total cost of care for the commercial population ($310 PMPM; 117% of the average) and an ASC Admission rate of 3.4 admissions per 1,000 person-months (133% of the commercial average). Other counties, such as Vernon and Columbia, had higher ASC Admissions (231% and 243% of the statewide commercial average) than Kenosha's percentage, but a lower total cost of care. While Vernon County's admissions were particularly high, its average cost was 96 percent of the statewide commercial average.

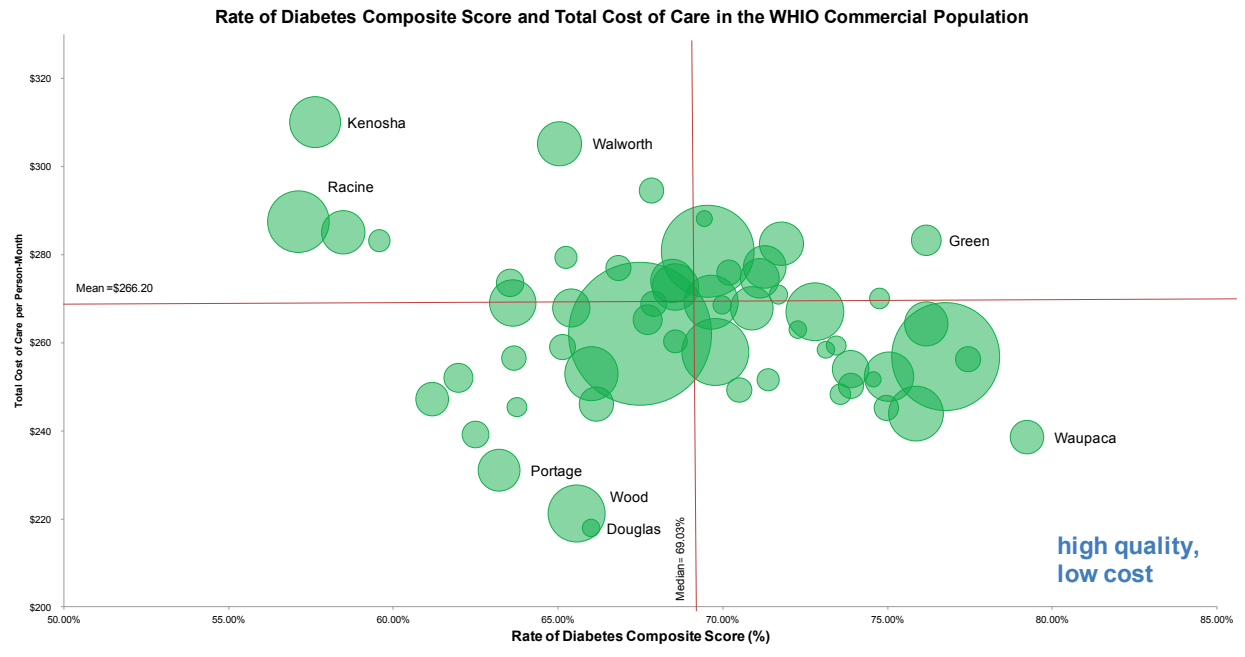
[44] Counties determined to have a small sample size are those for which N ≤ 100 for the HEDIS diabetes composite measure. Two counties in the Medicaid population were below this threshold (Kewaunee and Pierce). Additionally, 8 counties had an N between 100–150.

Exhibit 33. ASC Admissions and Total Cost for WHIO Commercial



The commercial WHIO population had a lower mean cost compared to the Medicare population ($263 PMPM; 41% of the Medicare cost) with a comparatively higher median composite score for the HEDIS diabetes measures (69% of the diabetics in the commercial WHIO population received all three tests, which was 118% of the median Medicare score). Twenty counties fell into the "high quality, low cost" quadrant, as defined by having total cost and HEDIS composite score below and above the population mean and median (Exhibit 34). Waupaca County in particular stands out in the high quality and low cost range with the fourth lowest cost by county—89 percent of the commercial average—and the highest composite score (79%) of all 71 counties (114% of the statewide average). Douglas and Wood counties both had lower costs than the payer mean (82% and 83% of the mean, respectively) and while their scores for the diabetes composite measure were lower than the average, they were still 96 percent of the commercial median.

Exhibit 34. HEDIS Composite and Total Cost for WHIO Commercial[45]



Consistency in Value of Care across Counties

Assessing the Value of Care across Counties and Payers Using the ASC Inpatient Admissions Measure

Exhibit 35 displays counties that are consistently low-quality and high-cost across multiple payers. Fourteen counties were found to be low-quality, high-cost across two or more payers when using statewide averages in ASC inpatient admissions and total cost indicators. Kenosha was the only low-cost county across three or more payers (Medicare, commercial, and Medicaid).

Exhibit 35. Low Quality, High Cost Counties for ASC Inpatient Admissions[46]

Low Quality, High Cost Counties in WHIO *(Counties qualifying as high quality and low cost across three or more payers based on ASC inpatient admissions and total cost)*				
	Medicare	**Commercial**	**Medicaid**	**Medicare Advantage**
WI, Adams	X	X		
WI, Barron	X	X		
WI, Chippewa		X		X
WI, Clark	X			X
WI, Columbia		X		X
WI, Kenosha	X	X	X	

[45] Counties determined to have a small sample size are those for which N ≤ 100 for the HEDIS diabetes composite measure. Four counties in the commercial population were below this threshold (Douglas, Dunn, Pierce and Polk). Additionally, seven counties had an N between 100 - 150.

Low Quality, High Cost Counties in WHIO *(Counties qualifying as high quality and low cost across three or more payers based on ASC inpatient admissions and total cost)*				
	Medicare	**Commercial**	**Medicaid**	**Medicare Advantage**
WI, Marinette		X	X	
WI, Milwaukee	X		X	
WI, Oneida	X			X
WI, Polk	X	X		
WI, Racine		X	X	
WI, Richland		X		X
WI, St. Croix	X	X		
WI, Walworth	X		X	

Conversely, Exhibit 36 displays counties that are consistently high-quality, low-cost across multiple payers using the same measurement criteria. Thirty of the counties in Wisconsin were high-quality, low-cost counties across more than one payer. Thirteen counties in Wisconsin (Door, Dunn, Fond du Lac, Green, Jackson, Jefferson, La Crosse, Outagamie, Portage, Price, Taylor, Waupaca, and Winnebago) were recognized as high-value counties across three or more payers. La Cross, Outagamie, Portage and Winnebago were recognized as high-value counties across all four payers.

Exhibit 36. High Quality, Low Cost Counties for ASC Inpatient Admissions

High Quality, Low Cost Counties in WHIO *(Counties qualifying as high quality and low cost across three or more payers based on ASC inpatient admissions and total cost)*				
	Medicare	**Commercial**	**Medicaid**	**Medicare Advantage**
WI, Barron			X	X
WI, Brown	X			X
WI, Calumet		X		X
WI, Dane	X	X		
WI, Door		X	X	X
WI, Douglas		X		X
WI, Dunn	X	X	X	
WI, Eau Claire		X	X	
WI, Fond du Lac	X		X	X
WI, Grant	X		X	
WI, Green	X		X	X
WI, Green Lake			X	X
WI, Jackson	X	X	X	
WI, Jefferson	X		X	X

High Quality, Low Cost Counties in WHIO *(Counties qualifying as high quality and low cost across three or more payers based on ASC inpatient admissions and total cost)*				
	Medicare	**Commercial**	**Medicaid**	**Medicare Advantage**
WI, Kewaunee	X	X		
WI, La Crosse	X	X	X	X
WI, Monroe		X	X	
WI, Outagamie	X	X	X	X
WI, Ozaukee	X		X	
WI, Portage	X	X	X	X
WI, Price		X	X	X
WI, Shawano	X			X
WI, Sheboygan	X			X
WI, Taylor		X	X	X
WI, Trempealeau	X		X	
WI, Vernon	X		X	
WI, Vilas	X	X		
WI, Waupaca	X	X	X	
WI, Waushara		X		X
WI, Winnebago	X	X	X	X

Assessing the Value of Care across Counties and Payers using a Composite Diabetes Measure
Exhibit 37 displays the 10 counties in Wisconsin that are consistently low-quality and high-cost across three or more payers when using the weighted statewide averages for the composite diabetes measure and total cost of care to assess value. Kenosha, Racine and Walworth counties were identified as low-quality, high-cost counties across all payers.

Exhibit 37: Consistently Low-Quality, High-Cost Counties in WHIO Based on Receipt of Diabetes Testing

Consistently Low-Quality, High-Cost Counties in WHIO Based on Receipt of Diabetes Testing				
	Medicare	**Commercial**	**Medicaid**	**Medicare Advantage**
WI, Adams	X	X		
WI, Chippewa		X		X
WI, Columbia		X		X
WI, Kenosha	X	X	X	X
WI, Manitowoc	X			X
WI, Marinette	X	X		
WI, Racine	X	X	X	X
WI, Richland		X		X

WI, Walworth	X	X	X	X
WI, Waukesha			X	X

Conversely, as Exhibit 38 shows, 26 counties in Wisconsin are providing high-value health care across multiple payers based on receipt of diabetes testing. Five counties (Dodge, La Crosse, Outagamie, Shawano, and Waupaca) were identified as high-value across three or more payers. Outagamie County provided high-quality, low-cost care across all payers.

Exhibit 38: Consistently High-Quality, Low-Cost Counties in WHIO Based on Receipt of Diabetes Testing

Consistently High-Quality, Low-Cost Counties in WHIO Based on Receipt of Diabetes Testing				
	Medicare	Commercial	Medicaid	Medicare Advantage
WI, Ashland	X		X	
WI, Brown	X			X
WI, Calumet		X		X
WI, Chippewa	X		X	
WI, Dane	X	X		
WI, Dodge	X		X	X
WI, Door		X	X	
WI, Dunn		X	X	
WI, Jackson		X		X
WI, Jefferson	X		X	
WI, La Crosse	X	X		X
WI, Marathon	X		X	
WI, Monroe		X		X
WI, Outagamie	X	X	X	X
WI, Pierce		X	X	
WI, Portage	X		X	
WI, Price	X		X	
WI, Richland	X		X	
WI, Rock		X		X
WI, Shawano	X	X		
WI, Sheboygan	X			X
WI, Trempealeau	X	X		
WI, Waupaca	X	X		X
WI, Waushara		X	X	
WI, Winnebago		X		X
WI, Wood	X		X	

Assessing the Value of Care across Multiple Quality Measures
The ASC inpatient admission measure and the composite diabetes measure were then examined in tandem to determine whether counties were providing high-quality health care at a low cost across both measures relative to total cost. Five counties fell into the low-value quadrant using both the ASC admissions and diabetes testing criteria (Adams, Chippewa, Columbia, Kenosha, and Marinette). Eighteen counties were recognized for having higher than average cost and quality for both measures (Brown, Calumet, Dane, Door, Dunn, Jackson, Jefferson, La Crosse, Monroe, Outagamie, Portage, Price, Shawano, Sheboygan, Trempealeau, Waupaca, Waushara, and Winnebago). These findings highlight the importance of measuring value not across geographies and payers and also across multiple measures.

Impact of Risk Adjustment on WHIO Results by Payer
As noted earlier, the above analysis on baseline costs and utilization was conducted using the highest tier of risk adjustment in each payer population. Separate analysis was undertaken in order to assess the relative impact of different tiers of risk adjustment on results.

As summarized in Exhibit 39, the impact of risk adjustment on the WHIO results can be determined by looking at the minimum, maximum, and median absolute differences in the least and most sophisticated risk adjustment methods (i.e., the difference between unadjusted results and the tier 1 adjustment of age and sex and the difference between unadjusted results and either tier 2 ERG adjusters for the non-Medicare data or tier 3 HCC scores for the Medicare data). The minimum and maximum absolute differences show the smallest and largest differences associated with each risk adjustment methodology across the 71 counties. The median absolute difference between different levels of risk adjustment for each measure is also displayed. Exhibit 30 also shows the absolute percent differences for the minimum, maximum, and median for each payer. In the case of WHIO, risk adjustment made a greater difference relative to the crude results than in PCIP and Blueprint.

In general, higher tiers of risk adjustment made more of a difference in results than lower tiers with a few notable exceptions: inpatient acute care admissions and days for the Medicare population for both the minimum and median differences and median differences for inpatient ASC days and for the minimum percent differences in the commercial cost results.

Overall, risk adjustment made very little difference in median results with two exceptions: inpatient acute care (20% difference between tier 1 and crude) and ASC days (15% difference between tier 2 or 3 and crude). In the case of Medicaid inpatient acute care admissions and inpatient ASC days, the two levels of risk adjustment had no impact on the percent difference in the median results. Large differences in magnitude in the minimum and maximum percentage differences make clear that even despite Wisconsin's largely homogeneous population, risk adjustment can make a large difference in results for outliers.

Exhibit 39: Impact of Risk Adjustment on WHIO Results

		Inpatient acute care admits per 1,000		Inpatient acute care days per 1,000		ASC acute care admits per 1,000		Inpatient ASC days per 1,000		Total cost of care cost	
		Tier 1 vs. Crude	Tier 2 or 3 vs. crude	Tier 1 vs. Crude	Tier 2 or 3 vs. crude	Tier 1 vs. Crude	Tier 2 or 3 vs. crude	Tier 1 vs. Crude	Tier 2 or 3 vs. crude	Tier 1 vs. Crude	Tier 2 or 3 vs. crude

		Inpatient acute care admits per 1,000		Inpatient acute care days per 1,000		ASC acute care admits per 1,000		Inpatient ASC days per 1,000		Total cost of care cost	
		Tier 1 vs. Crude	Tier 2 or 3 vs. crude	Tier 1 vs. Crude	Tier 2 or 3 vs. crude	Tier 1 vs. Crude	Tier 2 or 3 vs. crude	Tier 1 vs. Crude	Tier 2 or 3 vs. crude	Tier 1 vs. Crude	Tier 2 or 3 vs. crude
COMMERCIAL	**Minimum**										
	abs diff	0.03	0.00	0.00	0.13	0.00	0.01	0.00	0.02	0.16	0.08
	% diff	7%	16%	10%	23%	10%	17%	14%	21%	12%	10%
	Median										
	abs diff	**0.98**	**3.71**	**4.29**	**16.74**	**0.04**	**0.17**	**0.19**	**0.57**	**6.68**	**12.10**
	% diff	**0%**	**5%**	**1%**	**6%**	**1%**	**5%**	**2%**	**6%**	**1%**	**3%**
	Maximum										
	abs diff	4.54	22.62	19.47	72.89	0.51	1.39	1.48	3.51	33.46	56.24
	% diff	8%	43%	7%	46%	4%	38%	6%	43%	5%	26%
MEDICAID	**Minimum**										
	abs diff	0.03	0.30	1.00	0.55	0.00	0.05	0.03	0.29	0.05	0.65
	% diff	7%	33%	6%	36%	13%	35%	15%	38%	7%	29%
	Median										
	abs diff	**0.03**	**13.31**	**15.17**	**79.54**	**0.73**	**1.44**	**2.92**	**5.76**	**14.18**	**47.22**
	% diff	**2%**	**2%**	**1%**	**2%**	**2%**	**4%**	**3%**	**3%**	**1%**	**2%**
	Maximum										
	abs diff	0.03	79.86	46.67	351.16	4.25	8.55	15.48	24.17	39.92	387.46
	% diff	10%	73%	7%	79%	23%	85%	31%	98%	10%	292%
MEDICARE ADVANTAGE	**Minimum**										
	abs diff	0.03	1.54	1.32	3.36	0.09	0.16	0.02	0.03	0.58	0.98
	% diff	5%	21%	5%	26%	10%	22%	12%	24%	4%	31%
	Median										
	abs diff	**0.03**	**22.24**	**21.47**	**105.28**	**1.67**	**3.25**	**6.54**	**12.28**	**15.37**	**76.00**
	% diff	**1%**	**7%**	**1%**	**7%**	**3%**	**8%**	**3%**	**8%**	**1%**	**5%**
	Maximum										
	abs diff	0.03	73.71	63.51	399.15	6.77	12.35	23.46	48.94	53.55	419.00
	% diff	7%	88%	7%	106%	17%	96%	19%	104%	5%	67%
MEDICARE	**Minimum**										
	abs diff	0.03	0.53	45.30	3.41	0.22	0.02	0.00	1.57	1.05	20.00
	% diff	20%	10%	35%	20%	19%	22%	2%	56%	6%	8%
	Median										
	abs diff	**0.03**	**17.07**	**342.26**	**70.54**	**2.35**	**4.15**	**0.84**	**55.57**	**68.24**	**246.82**
	% diff	**8%**	**5%**	**20%**	**4%**	**3%**	**5%**	**0%**	**15%**	**1%**	**3%**
	Maximum										
	abs diff	0.03	98.46	661.57	383.57	36.70	44.70	7.16	340.86	723.84	1111.83
	% diff	13%	34%	3%	16%	68%	82%	1%	146%	11%	19%

Cross-Site Medicare Analysis

This section describes the variation in Medicare spending, utilization, and quality between Blueprint, PCIP, and WHIO. Comparing Medicare data across the three sites allows sites to see how they fare against each other on a select set of quality metrics and against national averages. Standard prices were used in each site to adjust for geographic differences in price.

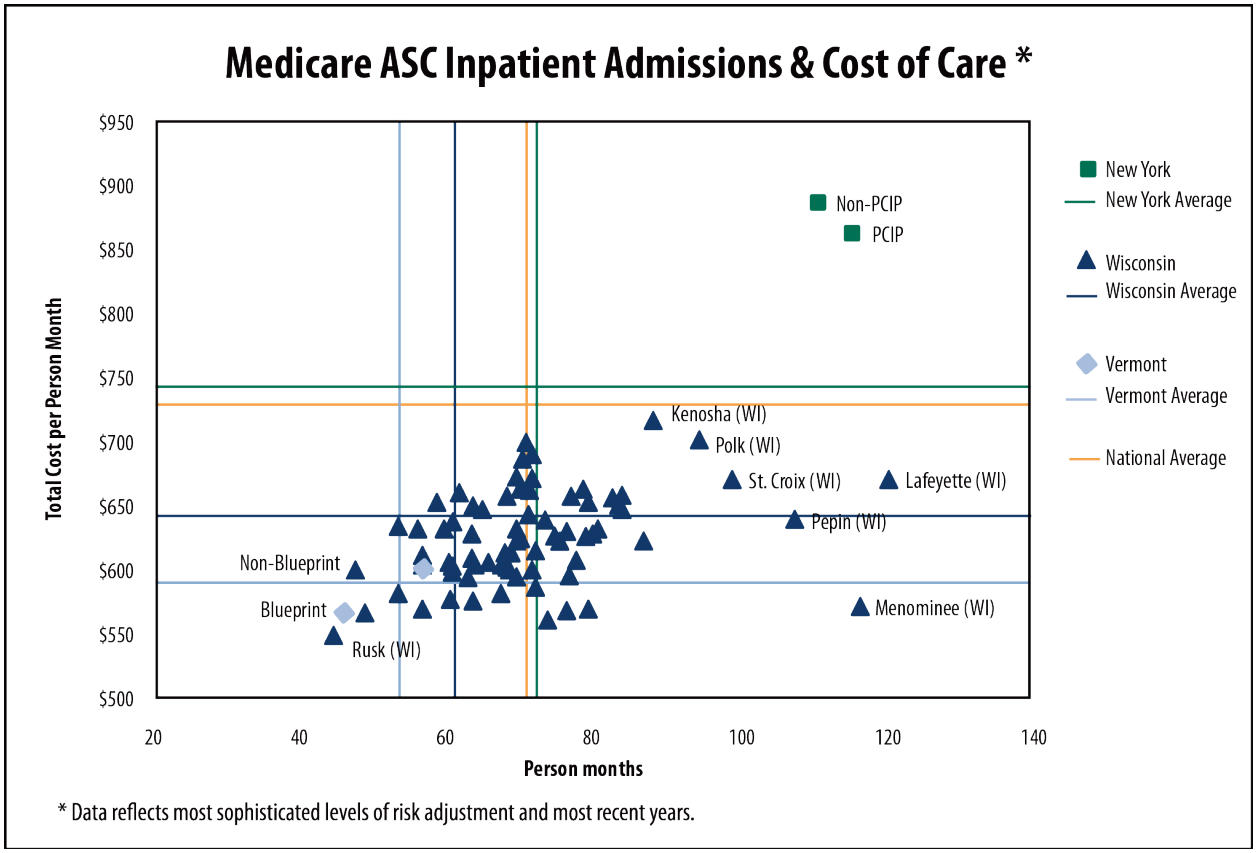
The national average total cost of care for the Medicare population is $725 PMPM. This is above the average cost for the states of Wisconsin ($639 PMPM) and Vermont ($601 PMPM) but below the average statewide cost in New York ($734 PMPM). While some site practices may not perform well when compared to state averages, in the context of national averages some sites might be doing better than it may seem. For example, while Kenosha County is clearly an outlier in cost in the WHIO Medicare populations, it is still performing under the nation's average total cost of care.

Inpatient ASC Admissions

Exhibit 40 below maps cost of care relative to ASC admissions for the Medicare populations for WHIO, Blueprint, and PCIP against national (orange lines) and statewide averages (green lines for New York, blue lines for Wisconsin, and light blue lines for Vermont). Of the three sites in this study, PCIP is the only site in a state with an average cost of care higher than the national average. The average cost of care for patients in New York was 107.2 percent of the national average, while Vermont and Wisconsin were 77.3 percent and 81.1 percent of the national average, respectively. Moreover, patients in New York attributed to PCIP were 158.4 percent of the national average in ASC admissions and 118.8 percent of the national average in cost. Compared to the New York state average, PCIP patients were 148 percent of the statewide average in ASC admissions and 117.2 percent of the New York average in cost.

While both the PCIP comparison and intervention groups had higher costs compared to the other sites, two counties in Wisconsin had slightly higher rates of ASC inpatient admission per 1,000 person-months—Lafayette and Menominee. However, both counties had lower total cost compared to the PCIP comparison and intervention groups.

Exhibit 40. Cross-site Medicare ASC Admissions and Cost of Care

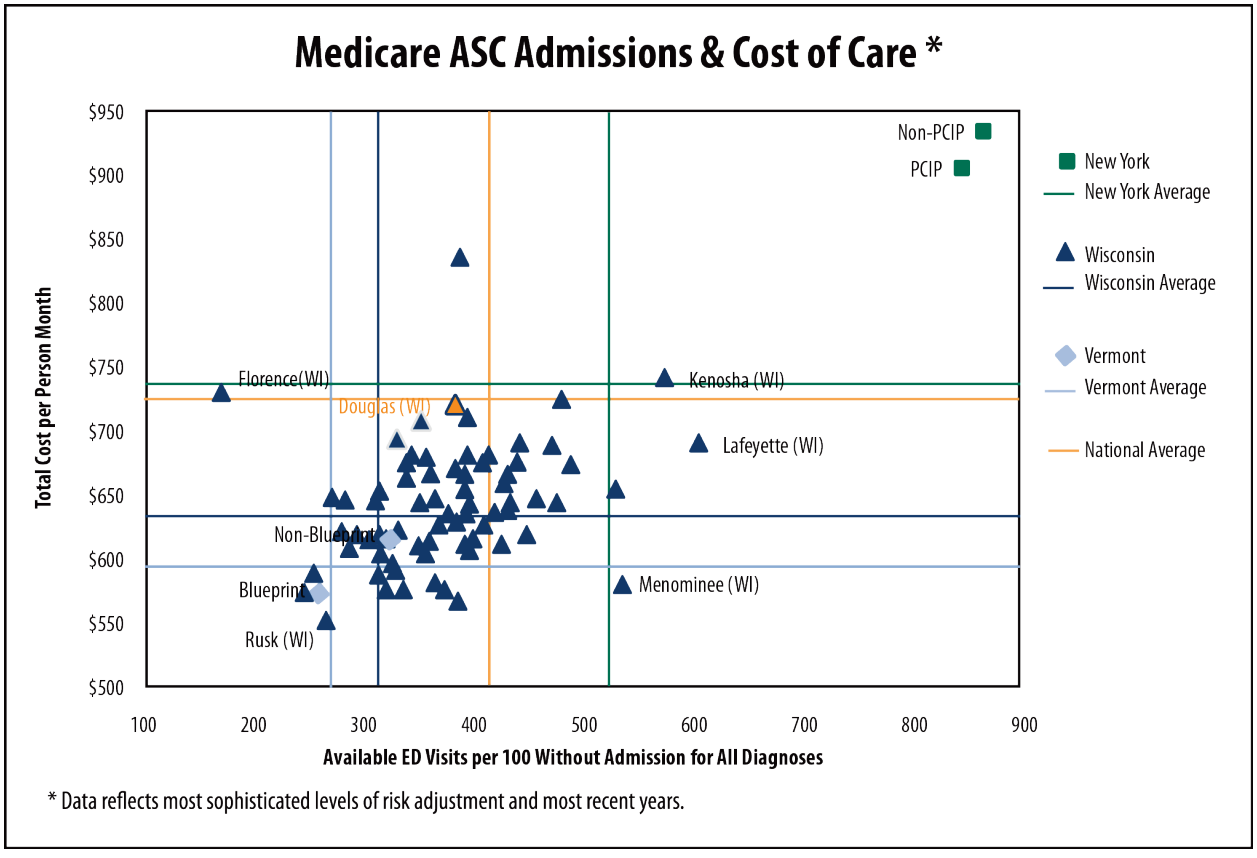


Inpatient ASC days

The national average for ASC inpatient days for the Medicare population is 418.6 days, which is above the Vermont and Wisconsin averages of 275.4 and 354.72 days, respectively. New York has the highest statewide average with 529.6 days per 100. Compared to national averages, PCIP participants had more than double the national average at 203.1 percent while those attributed to the comparison non-PCIP group were at 207.4 percent of the national average. Both Blueprint comparison and intervention groups appear to be performing best out of the three sites when compared to national averages (61.6% of the national average for ASC days in Blueprint and 77.6% in non-Blueprint and 8–16% percent below national averages in cost). However, when compared to the state average, the Blueprint and comparison groups did not fare as well. ASC days for Blueprint and non-Blueprint groups were 101.4 percent and 111 percent of the statewide average and 111 percent, respectively.

When measuring the value of care based on ASC inpatient days relative to the total cost per person-month, a number of the Wisconsin counties can be labeled as low-quality, high-cost compared to the state averages. However, when compared to the national averages for ASC inpatient days (72.6 days) and the average national cost of care for Medicare beneficiaries, many of the Wisconsin counties are performing well (see Exhibit 41). For example, Douglas County has higher ASC inpatient admissions (123% of the state average) and a higher cost (109% of the state average) than its peers in Wisconsin, but still qualifies as a high-quality, low-cost county when compared to national averages with ASC admissions 92 percent of the national average and a total cost 96 percent of the U.S. average cost of care PMPM.

Exhibit 41. Cross-site Medicare ASC Inpatient Days and Cost of Care



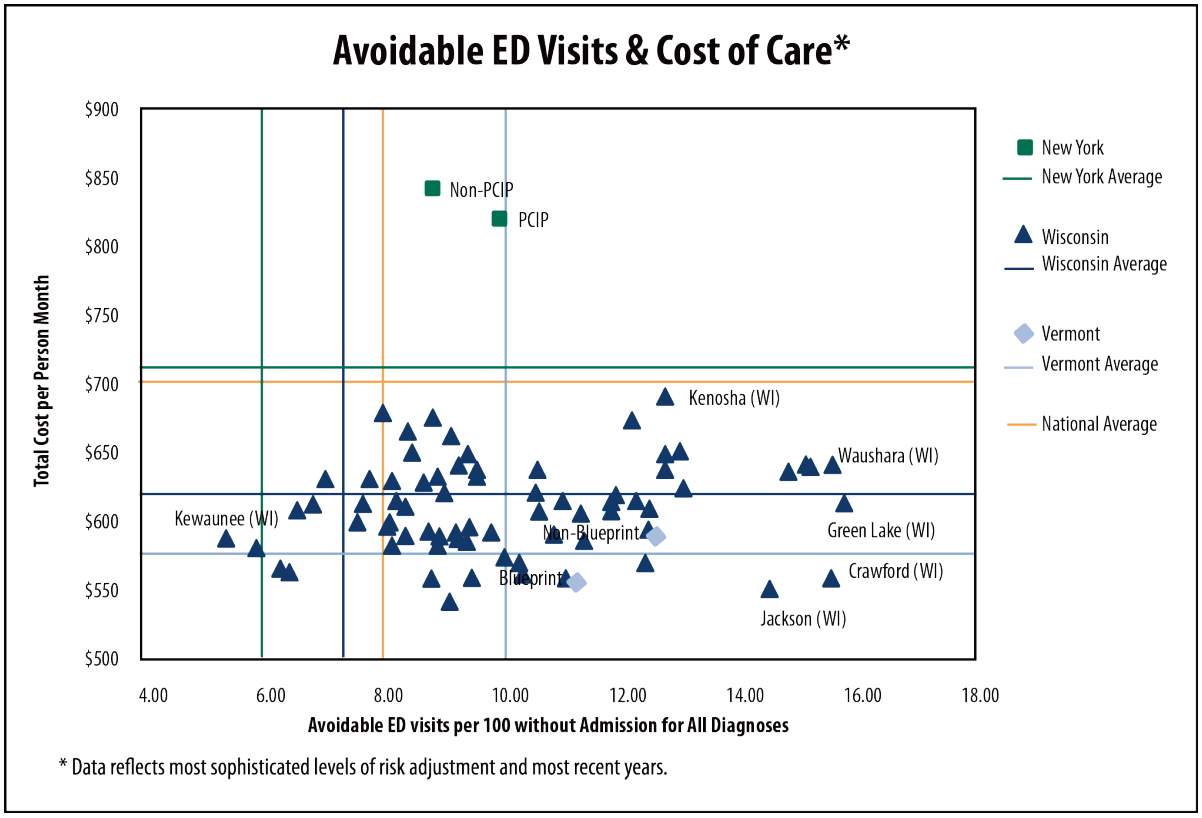
Avoidable ED Visits

Avoidable ED visits is the only statewide average among the suite of quality metrics examined in this project for which New York outperforms other states (Exhibit 42). The average number of avoidable ED visits in NY is 6.1 per 1000 person months, which is lower than the Wisconsin average of 9.81 per 1,000 person-months and the Blueprint average of 9.9 per 1000 person-months. New York is 75 percent of the national average of 8.1 avoidable ED visits per 1,000 person months, while Wisconsin and Vermont are 121 percent and 122 percent of the national average, respectively.

Similarly, the PCIP comparison and intervention populations also had lower avoidable ED visits than a number of the other practice sites in this project. While this could reflect better access to appropriate primary care, however, this could be an artifact of the way the measure is defined. Because ED visits that resulted in an inpatient admission are excluded, more severe ED visits might not be captured in this particular measure. As observed in Exhibits 40 and 41, New York tends to have higher rates of avoidable inpatient admissions, and those inpatient admissions also tend to result in longer lengths of stays. It should be noted that the PCIP and non-PCIP groups still had more avoidable ED visits than the national average (PCIP was 129.6% of the national average and the non-PCIP group was 109.9% of the U.S. average).

Blueprint and non-Blueprint groups also had more avoidable ED visits than the U.S. average at 135.8 percent and 160.5 percent of the national average, respectively. Of all project sites, Green Lake County in Wisconsin had the highest number of avoidable ED visits at 197.5 percent, almost double the national average.

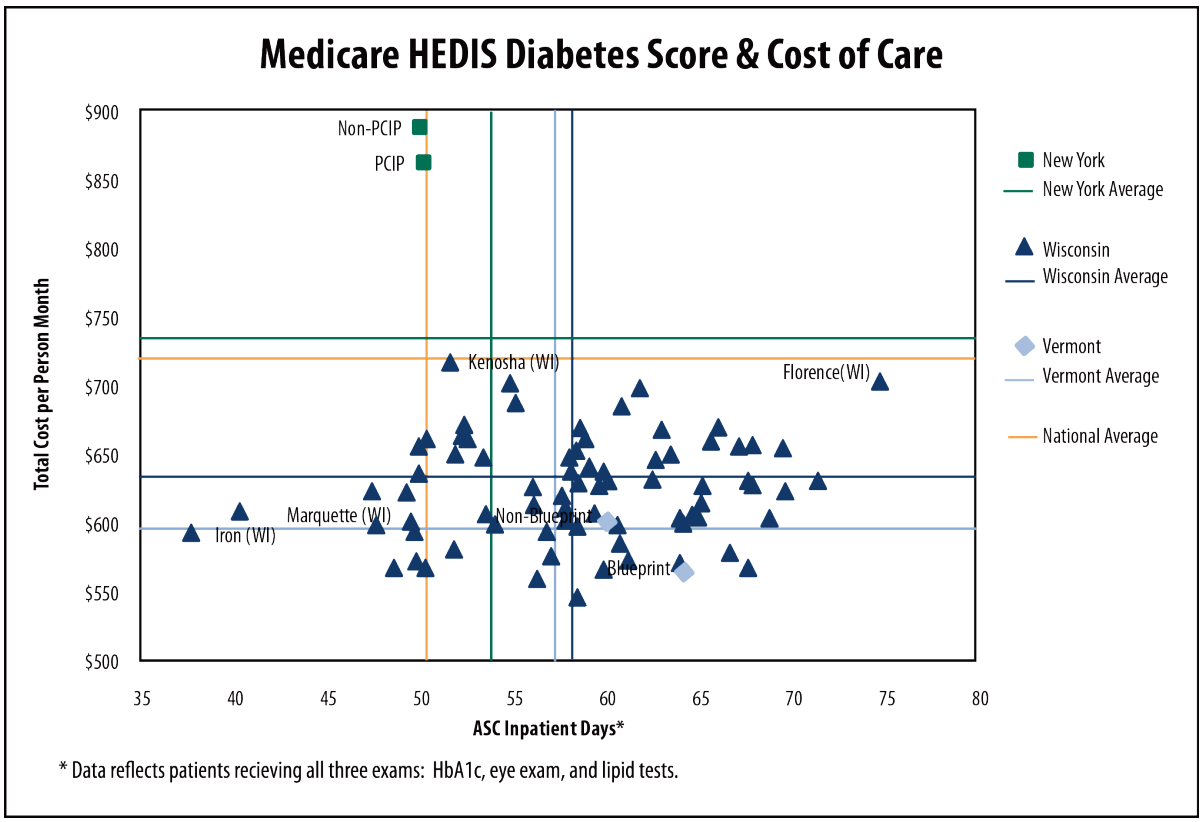
Exhibit 42. Cross-site Medicare Avoidable ED Visits and Cost of Care



Quality of Diabetes Care

For both Wisconsin and Vermont, the statewide average for the diabetes composite score was between 55 and - 60 percent (Exhibit 43). All of the practice sites outside of New York fell below the New York state average in cost ($734 PMPM) and each of the Blueprint sites, and many of the WHIO counties, outperformed New York's statewide average of 58.3 percent receipt of all three HEDIS diabetes tests. Both the Blueprint and non-Blueprint populations outperformed Vermont's statewide averages for total cost ($601) and the HEDIS diabetes score (57.2%). The Blueprint and non-Blueprint groups, along with most Wisconsin counties performed above the national average composite score (50.4%). The PCIP and non-PCIP slights fell slightly below the U.S. average.

Exhibit 43. Cross-site Medicare HEDIS Diabetes Testing Score and Cost of Care



Summary of Findings

Vermont

Differences in population characteristics at baseline: Differences in the Blueprint participants and the non-Blueprint populations for commercial and Medicare data were minimal. Overall, the Medicare Blueprint participants had similar health and demographic status when compared to the non-Blueprint populations. Exceptions to this were slightly lower mortality rates in the Blueprint Medicare participants. Blueprint participants also lived in zip codes in which residents were more educated. Blueprint participants also had higher rates of diabetes. However, differences in the overall HCC score were not statistically significant.

Overall, the commercial populations in Vermont were also similar. While data on race is not included in the Vermont database, the age and sex of both the Blueprint and non-Blueprint commercial populations were approximately the same in both sites. Members of the commercial Blueprint population in St. Johnbsury had slightly higher rates of diabetes than those not in Blueprint and lower rates of asthma. In Burlington, Blueprint members had slightly higher rates of coronary heart disease.

Cost and utilization differences at baseline: In Vermont, Blueprint and non-Blueprint populations were roughly similar at baseline—particularly in the commercial population—with a few notable exceptions. Of 24 cost and utilization measures, only six to seven were statistically significant in the St. Johnsbury and Burlington commercial populations. In the Medicare population, differences in 15 out of 27 cost and utilization measures were statistically significant.

Where differences were statistically significant, however, different patterns emerged among the Medicare and commercial populations. In the Burlington site, the Blueprint commercial population had higher cost and utilization rates than their non-Blueprint counterparts on six of seven statistically significant results. However, in the overall Medicare and commercial populations in St. Johnsbury, spending and utilization tended to be lower in the Blueprint population relative to the comparison group. Where utilization was higher among the Blueprint populations, they tended to be higher for primary and secondary preventive services (e.g., primary care visits and visits to specialists) and lower for ambulatory-sensitive care measures like avoidable ED visits and hospitalizations.

Quality differences at baseline: Using the composite diabetes test result as a proxy for overall care quality, the receipt of clinically-recommended screening tests was higher among the Blueprint Medicare and commercial populations than their non-Blueprint counterparts, which may be partly due to a previous intervention that aimed to improve quality of care for patients with diabetes that was already initiated prior to the start of Blueprint. Moving forward, keeping in mind that practices join Blueprint voluntarily and therefore differences could be in the practices themselves, rather than the intervention, is important.

PCIP

Differences in population characteristics at baseline: Overall, the Medicare PCIP participants were similar to the comparison group. With the exception of gender, the Medicare PCIP group differed in statistically significant ways from those not enrolled in PCIP, however, the magnitude of those differences was small. The PCIP participant population tended to be slightly younger,[47] more racially diverse as recorded in claims records, lived in zip codes in which residents were less educated, and in a lower zip income bracket than the non-PCIP population. A greater proportion of PCIP patients also had diabetes and were initially enrolled in Medicare for a disability. However, PCIP participants also tended to have lower HCC scores, and a smaller proportion had CHF or COPD compared to those not enrolled in the program.

Similarly, Medicaid patients enrolled in the PCIP program differed slightly from those who were not enrolled in the program. Overall, PCIP Medicaid patients tended to be older, lived in zip codes in which residents were less educated, and had a lower zip income. A greater proportion of PCIP participants were recorded as Hispanic and female in claims records, while a smaller proportion was recorded as black, compared to their non-PCIP Medicaid counterparts. PCIP Medicaid beneficiaries appeared to be slightly healthier than those not participating in the program, based on cases of CHF and COPD, but did have slightly higher rates of diabetes. A greater proportion of Medicaid PCIP patients were also enrolled in managed care compared to their non-PCIP Medicaid counterparts.

Cost and utilization differences at baseline: Compared to Vermont, the differences between PCIP and non-PCIP populations at baseline were much more pronounced. Differences were statistically significant for 25 of 39 cost and utilization measures for the Medicare population, and for 23 out of 32 cost and utilization measures for the Medicaid population. However, it is important to note that while these results were statistically significant, the magnitude of the differences between the PCIP and non-PCIP groups were small.

[47] Medicare beneficiaries assigned to practices in the PCIP and non-PCIP groups were all over the age of 64, but those enrolled in PCIP tended to be younger than those in the non-PCIP group.

Baseline data showed that the Medicare PCIP population generally had lower utilization and cost less overall than those not enrolled in the program. However, patients participating in PCIP were more likely to have higher rates of ED use, including avoidable ED visits. The same held true for the PCIP Medicaid population, which had lower rates of cost and utilization than their counterparts with the exceptions of allowed charges for E&M, procedures and tests, and outpatient visits to primary care providers. Given the demographic profile of the PCIP population described above, assisting Medicare patients to better navigate the health care system and improving care coordination to prevent high cost incidents like avoidable ED visits might be an important area for PCIP to focus their future efforts on as health care reform progresses. Within the Medicaid population, PCIP might look into what is driving the higher allowed charges for E&M, procedures and tests.

Quality differences at baseline: Overall, both Medicare and Medicaid patients in both the PCIP and non-PCIP groups appear to be receiving similar levels of care quality, however, those enrolled in the PCIP group are receiving that care at a lower cost at baseline. Moving forward, keeping in mind that practices join PCIP voluntarily and therefore differences could be in the practices themselves, rather than the intervention, is also important. Multivariate logistic regression results suggest that sex and race have important independent effects on the likelihood of receiving recommended care, with women and recorded blacks being less likely to receive the care they need in the Medicare population. Moving forward, reducing such racial disparities might be an important dimension of the reform initiatives in New York City.

WHIO

Because the objective of the WHIO assessment was to identify targets for future reform efforts, analysis focused on identifying those counties that appear to be delivering either consistently high-value care (high-quality and low-cost) or consistently low-value care (low-quality and high-cost) across multiple payers when compared to statewide averages. Allowing counties to learn best practices from one another can be a useful tool for scaling up reform efforts. As such, future resources could be allocated towards learning how the high-value counties have achieved success and helping lower performing counties implement those best practices.

Eighteen counties (Brown, Calumet, Dane, Door, Dunn, Jackson, Jefferson, La Crosse, Monroe, Outagamie, Portage, Price, Shawano, Sheboygan, Trempealeau, Waupaca, Waushara, and Winnebago) were identified as providing health care that was better quality and lower cost than the statewide payer average across two or more payers according to two different quality metrics.[48] These counties tend to cluster in the mid-eastern region of the state, along with a small group of counties along the middle of the western border. Based on clinical care factors, five of these counties ranked in the top 10 (La Crosse, Outagamie, Dane, Winnebago, and Waupaca).[49] Jefferson and Brown counties also qualified in the top 20. Outagamie County was recognized as a high-value county across all payers for both of the quality metrics used in this study. La Crosse County is also notable for performing as a high-value county across all payers in the ASC admissions measure and across three payers (all but Medicaid) in the diabetes composite measure. These are the counties that Wisconsin might seek to learn best practices from and

[48] Value of care is measured by looking at ASC admissions and the composite HEDIS diabetes measure, relative to total cost of care (based on standard prices) for each payer. Counties are then determined to be low-quality, high-cost by comparing county results to the statewide averages.

[49] County Health Rankings: Mobilizing Action Toward Community Health. 2011 Wisconsin. http://www.countyhealthrankings.org/sites/default/files/states/CHR2011_WI_0.pdf. Clinical care factors included measures of access to care like uninsured rates and quality of care measures like preventable hospital stays, diabetes screening, and mammography screening.

share those methods with other counties, should further analysis across several other measures demonstrate a consistent pattern.

Five counties (Adams, Chippewa, Columbia, Kenosha, and Marinette) were identified as providing lower quality and higher cost health care than the statewide payer average across two or more payers according to two different quality metrics. Kenosha County had higher than average cost and lower than average diabetes care quality across three payers using the ASC measure and across all four payers in testing for diabetes. These findings are also largely consistent with the latest county health rankings.[50] Adams County was ranked as the second lowest county in health care outcomes, while Marinette and Racine counties were ranked as numbers 63 and 59, respectively. Practices in these counties might be high-priority areas for the state of Wisconsin to improve health care quality and costs through reform efforts.

However, caution should be exercised when interpreting these results because tests of statistical significance were not performed and because the population sizes of some counties and payer populations may be too small to yield stable results.

Analyzing Medicare Cost, Quality and Utilization across Sites

Benchmarking site data to national averages can help sites understand the context of the health care environment in which they are implementing their reforms and potentially support their efforts to identify areas for improvement. For example, while PCIP Medicare participants tended to have lower total costs of care than those not participating in PCIP, both comparison and intervention groups were well above the national and state-level averages for total cost. Both groups were also above state and national averages for select quality measures (ASC inpatient admissions, ASC inpatient days, and avoidable ED visits).

In the case of Vermont, by contrast, the Medicare populations for both the Blueprint and non-Blueprint groups are low-cost and high-quality when compared to national averages. Blueprint had lower rates of ASC inpatient admissions and days than the non-Blueprint group and fared better than the statewide average in these two measures.

In the case of Wisconsin, a number of counties that were low-quality, high-cost in the context of statewide Medicare averages were actually high-value counties when compared to national averages. This was particularly true in measures of ASC inpatient admissions and inpatient days. Only five counties in Wisconsin did not fall into the high-quality, low-cost quadrant when compared to national averages for ASC admission and total cost. Moreover, all of the counties in Wisconsin provide care to Medicare beneficiaries at a lower cost than national benchmarks. When the same comparisons were drawn for the Wisconsin Medicare Advantage population, all Wisconsin counties fell into the high-quality, high-cost quadrant for ASC admissions and total cost of care.

While Wisconsin should aim to reduce variation in quality and spending statewide, it is particularly important to focus on counties that are far from statewide and national averages.

[50] Retrieved December 13, 2011, from http://www.countyhealthrankings.org/wisconsin/overall-rankings. Health outcomes include measures of mortality like premature death and morbidly like low birth weight and poor self-reported health status. Health factors include behavioral influences on health like tobacco use and activity level.

Chapter 4
Discussion

Chapter 4: Discussion

In this chapter, challenges encountered and lessons learned during the course of this project are discussed in the context of broader health reform considerations.

Building an Evaluation Template for Community-Based Reform Initiatives: Challenges and Lessons Learned

Policy makers have long recognized that, "all health care is local." However, policy makers are only beginning to realize that achieving substantial and lasting impacts in payment and delivery reform will require engaging whole communities in comprehensive reform efforts, rather than the application of individual interventions among isolated provider groups or single health plans. The emergence of community-based initiatives like the Beacon Community Program and AF4Q reflect this growing recognition. However, community-based initiatives—especially those that involve data from multiple payers—raise novel implementation and evaluation challenges.

In the context of the current health care environment, this project can provide timely and relevant insight into the role that community-based, multi-payer health care initiatives can play in improving population health and how to assess such interventions. Extrapolating lessons learned from the hard technical issues encountered during work with the three sites can provide valuable insight into the methodological options and best practices for measuring the impact of reform interventions in local communities and, therein, inform a broader template for evaluation of the impacts of various reforms across other multi-payer sites.

Aligning a Common Template with Specific Health Goals for Different Sites and Populations

To the extent possible, this project aimed to create a standard template that could be agreed upon and used across the three communities, to facilitate high-level cross-site analysis. Developing a common framework that can be used by the three diverse sites in this project and across multiple payers also increases the likelihood that the template could also be used in the evaluation of other community-based initiatives.

However, in order to provide the most appropriate evaluation framework for each site and population—and in light of data and resource constraints—in some instances it was necessary to tailor measure specifications and methods to take into account considerations unique to those communities or populations within those communities.

Establishing a Baseline

In the case of Blueprint, baselines were defined differently for the Medicare and commercial populations. While Onpoint used staggered start-up dates for the commercial data to take into account the dates that interventions were rolled out in each site, Medicare used calendar year data. All available Medicare data predated the start of the interventions, so was still considered baseline. Differences in the sample size also led to some variation in the years of data being used. For example, because Dartmouth only had a 20 percent sample size for the Medicare data, data from 2005 to 2007 was time adjusted to produce a single rate.

Approaches to Creating Pools for Comparison

Different approaches were also taken when determining population pools for comparison. Weighted counties were used to create a comparison population for the Medicare population in Vermont, while Onpoint chose to use propensity score matching for the commercial population. While both approaches

removed most of the variation between the cases and matched controls, the propensity score matching method was also the more time-intensive of the two approaches. When selecting methods for evaluating an intervention, sites should consider the resources and time that they have available.

Standardized Metrics
Generally, differences that existed around the measures used between sites and payers were due to a lack of the necessary resources and funding rather than limitations around the data. A few exceptions in the New York Medicaid data. Since New York Medicaid data does not show ED encounters that result in hospitalizations, only encounters and costs were included for the ED measures. New York Medicaid data also does not break diabetes testing down by HbA1c tests and blood lipid tests, so only eye exams were included.

Measure specifications varied most in the expenditure categories. While WHIO and PCIP relied on standard prices, for example, Vermont had access to allowed charges. The measure specifications for standard payments also varied slightly, with pharmacy data included in the WHIO commercial and Medicaid calculation of total costs but not included in the Medicare calculation of total cost. Some items (e.g., home health and skilled nursing facilities) could not be converted into standard payments that would allow for comparisons across geographies because they are billed using multiple metrics in New York (e.g., per hour, per month, capitated, etc.).

Risk Adjusting for Different Populations
Risk adjustment also varied slightly among the various communities and payer groups. For example, the non-Medicare populations in WHIO and Blueprint used ERG scores instead of HCC scores since HCC scores are not as appropriate for a population under 65. PCIP included an indicator for managed care in the third tier of risk adjustment to take into account the large percentage of the population that were enrolled in managed care. The other major difference in risk adjustment methods was in the Blueprint commercial population, which did not include adjusters for race because race information is not available in the database. The Blueprint commercial population was the only population for which multiple tiers of risk adjustment were not applied. Across all other sites and payers, multiple tiers of risk adjustment were used. When evaluating community-based interventions, multiple tiers of risk adjustment can be useful to allow sites to pick and choose which level of risk adjustment is most appropriate for a specific measure and payer.

Allowing for Flexibility over Time
Additionally, it is important to note that even after carefully vetting different methodologies, approaches are subject to change as new data and technology become available. For example, the Blueprint Program has already begun to make adjustments to their evaluation methods. In the future, time periods for reporting will shift to calendar years as opposed to the staggered dates initially assigned to the Blueprint commercial population. Episode risk groups will be replaced by clinical risk groups (CRGs) for health status matching in the non-Medicare populations. Length of enrollment may also be integrated into the matching criteria. While the initial evaluation only included populations ages 18–64 in the commercial group, children will be included in future evaluations. A new intervention site (Barre) will also be represented in the baseline data for the commercial population. Lastly, Blueprint is reviewing their attribution methods and may consider different approaches to attribution in the future.

Identifying and Selecting Metrics

The Need for Holistic Analysis of Cost, Quality, and Utilization Metrics across Payers

This project demonstrated how important it is to look across various types measures and assess them holistically to assess effectiveness of interventions. Looking at cost alone in the Wisconsin commercial data, for example, Douglas and Wood counties both appeared to be the two lowest-cost counties with $218 and $221 PMPM, respectively. However, when cost was looked at alongside ASC inpatient admissions, Douglas County emerged as a high-value (low-cost and also high-quality) county, while Wood County had more ASC inpatient admissions than the statewide average. In other words, while Wood County appeared to be efficient when looking at cost alone, its ASC inpatient admissions rate suggested that low spending may be coming at the expense of quality.

Another example of this was in the WHIO Medicare population. Iron County was one of the counties in Wisconsin with the lowest total cost of care for Medicare beneficiaries. However, when cost and HEDIS diabetes measures were examined together, Iron County also had the lowest rate of diabetic Medicare patients receiving HbA1c, eye exams and blood lipid tests. Only 38 percent of Medicare beneficiaries received all three exams in Wisconsin while the statewide average for Medicare was 59 percent.

The Need for Multi-Payer Metrics

In addition to looking at different measures together, this project demonstrated why it is so important to consider the value of health care across payers. In Wisconsin, for example, total cost for those in Douglas County with commercial, Medicaid and Medicare Advantage coverage was lower than the statewide average for those payers. In the Medicare population, however, standard payments were higher in Douglas County than in all but two other counties in Wisconsin.

While the reasons for these differences are varied and complex and may involve elements of cost-shifting, this analysis demonstrates why evaluating total cost of care over multiple payers is so critical to accurately assessing health care value and system-wide effects on care cost and quality.

The Need for a Common Set of Consistently Defined Measures

This ability to compare across different analytic groups—whether the groups are health care providers, payers, counties, states, or whole interventions—underscores the importance of a common core set of consistently defined measures.

Consistent measures enable reliable equivalent comparisons, which will become increasingly important as health reform implementation progresses and health care professionals are held more accountable for the care they provide through performance-based payments. Providers will seek assurances that the metrics are valid and calculated reliably. Measures that can be aggregated across data sources to provide a comprehensive view of entire provider caseloads, rather than slices of performance data by payer population, will also be more meaningful to providers and to consumers for quality improvement purposes and informed consumer decision-making.

When used to compare a variety of promising payment and delivery reform efforts, common core measures facilitate the rapid evaluation of their comparative effectiveness which, in turn, enables decision-makers to make evidence-based decisions on which models are most appropriate for widespread dissemination. Even when interventions target specific patient populations, such as the elderly Medicare population, a consistent set of multi-payer metrics can be very valuable for the evaluation of spillover effects and unintended effects like cost-shifting.

Developing Consistent Risk Adjustment Methods

Aside from prices and utilization, risk adjustment can also be a major driver of variation in health spending (and utilization). Health care cost and use has been shown to vary significantly by age, gender, race, income, insurance status, prior health history, and many other factors. Adjusting for these differences is critical to understand whether spending variation in simply due to differences in the underlying population or is actually related to the performance of the various providers. Recognizing this, all cost and utilization measures calculated under this project were risk adjusted, though crude results were also provided.

Because measures were calculated in this project for extremely different populations (i.e., Medicare, Medicaid and commercial populations), developing a single risk adjustment method across all populations would have been extremely difficult. The same data were not available for risk adjustment across all payer sources. For instance, many payers do not keep track of all procedures and diagnosis codes for all services rendered.

Rather than attempt to develop a single risk adjustment methodology, the approach taken in this project was simply to agree that TCC measures should be adjusted for age, gender, other socio-demographic variables (e.g., race and income), and health status, leaving the particulars up to each site depending on data availability and population characteristics. As CMS and commercial payers undertake more value-based payment initiatives that are being driven by TCC and related measures, it is advisable to start converging towards more similar risk adjustment methods.

Challenges in Implementing a Consistent Total Cost Measure

Although total cost measures can be powerful policy tools and are widely used, no standard methods for such measure calculations currently exist. Efforts have, for the most part, focused on Medicare spending, but it is critical to also evaluate the spending patterns of the Medicaid and commercial populations also, for the reasons elaborated above. Evaluating commercial payer data in this manner is difficult because there is no single data standard. Developing these standards will be increasingly important as more multi-payer delivery reform initiatives are developed and implemented.
Although the concept of total cost measures is easy to understand and seems easy enough to operationalize, this report highlights the major issues that arose during the measure development process with the three sites as part of this project.

Data Variability and the Need for a Distributed Approach to Implementing Measure Specifications

Overcoming the wide variation in the way billing and utilization data is collected and categorized across payers, which stem largely from the fact that payers reimburse providers for specific services differently, was one of the most difficult problems in this project. For example, Medicare typically reimburses for inpatient hospital services on a discharge basis whereas many private payers do so on a per diem basis. Other issues stemmed from variation in the amount of time it takes different payers to process claims used to develop cost and utilization data, formats for housing the data, and general preferences for reporting and categorizing services.

The differing formats, elements, and quality of data from different payers made conducting analysis across payers challenging. Anticipating these challenges, this project relied on a distributed approach to data analysis, which minimized the transfer of raw patient-level data and enabled the data to remain closer to the source. Though this distributed approach required extensive coordination, this project has demonstrated that such an approach is not only viable but also advantageous. Not only is this a more privacy-protective approach, it facilitated the more timely analysis of the data because it enabled the

individuals most familiar with the intricacies of a particular data source to be responsible for cleaning and standardizing the data and running the measure specifications again that data.

Different Provider Identifiers
A major part of the work conducted with the sites was attributing patients to providers participating in delivery reform interventions. Being able to identify providers and their specialties across data sets during the attribution processes was critical. Unfortunately, many of the payers used different provider identification numbers and often reported unreliable specialty designations. This resulted in significant delays as additional provider information needed to be gathered. All attempts were made to use the NPI as the standard and it is recommended that all payers collect this information moving forward.

Differences in Benefit Design
Payers insuring different populations and often having different benefit packages also created difficulties when trying to standardize TCC measures. This was particularly problematic as some payers carved out certain services like physical therapy or prescription drugs. Other differences were due to variance in co-pay mechanisms.

It is possible to include an actuarial adjustment factor to control for many of the differences in co-pays. However, the impact on total cost due to variance in co-pays may be an effect of interest. In these cases, such an adjustment factor may not need to be included. Another consideration related to differences in co-pays was whether to report payer spending or total spending (including patient co-pays). Total spending was the preferred approach as it is a better indicator of total resources used and hence efficiency.

In order to deal with the carve-out issue, TCC measures were built up from component measures based on the major service categories of care delivered to patients. As part of this project, service categorizations typically included inpatient hospital, post-acute care, and ambulatory services. Prescription drug spending was generally excluded as this data was not available for most payers. Having the component cost estimates could allow for more standardized aggregations across payers with different service carve-outs.

Changes in Enrollment
As described above, one of the core aspects of TCC measures is that costs are calculated for a defined population. This is particularly critical as TCC measures are often reported on a total cost of care or PMPM basis in order to facilitate comparisons of populations with different sizes. For this reason, changes in plan enrollment, which occur frequently as enrollees move, change or lose jobs, pass away, and elect different plans, posed considerable challenges.

In general, adjustments should be made whenever possible to control for the fact that people may not have been enrolled for the entire timeframe of interest. For example, in situations where total cost is being measured for a given plan in a calendar year but where some enrollees were only enrolled for half the year, the costs for those enrollees could be adjusted for the fact that they cover a smaller timeframe. In many cases, those costs would be given a smaller weight (e.g., 0.5) when calculating the TCC measure standardized for population size to compensate for the fact that those enrollees represented less than full enrollment (only half of the year).

Reporting on a PMPM basis can avoid potential confusions due to differential weighting, as the denominator is simply the number of months or enrollment for all enrollees during the time frame of

interest. This works particularly well for health insurance because insurance coverage is often applied in monthly installments so there is no need to be concerned with partial month enrollment. Thus, there is no reason to apply additional weighting factors when reporting TCC measures on a PMPM basis. For payment purposes, PMPM TCC figures are also preferred since payments based on prospective budgeting are typically calculated and paid on a PMPM basis.

While reporting costs on a PMPM basis are advantageous for the reasons noted above, it may not be as intuitive as annual costs for some stakeholders. Therefore, considering the target audience when determining whether reporting a monthly or yearly figure is most appropriate, as it is often easier for many policy stakeholders to think in annual terms, is important.

Changes in enrollment are more problematic to deal with when trying to attribute patients to providers or delivery reforms for evaluation and payment purposes. One question that arose during this project was how long patients should be attributed to particular intervention providers before the providers are held accountable for their costs and utilization. Each of the interventions aimed at improving the way patients' overall health is coordinated and managed. This is difficult for providers to achieve if they only have a limited amount of time with the patients. For this reason, patients were required to be enrolled by their plan for at least nine months during the calendar year before being attributed to a provider.

The Lack of Good Price Data

Variation in TCC measures can be explained by several factors. One of the key drivers for spending variation is variation in pricing levels, and availability of this information can be extremely limited. Typically, this information exists for public payers who regulate reimbursement rates to providers. Hence their prices are fairly similar across all providers. Price levels are not as accessible for private payers who typically negotiate reimbursement rates with providers, which can lead to wide variation in prices for the same services even within a plan.

This is not always a problem, but it depends on the intended use of the information. People may be interested in knowing the total amount of money spent on health care regardless of what accounted for the spending levels. For most evaluation applications it will be important to control for pricing variation. The typical approach is to use a standardized pricing list to assume that the same price is charged for a service regardless of where the service was provided. In so doing, all variation in spending directly attributed to pricing is eliminated.

In some cases, actual spending data were not available due to sensitivity related to potentially revealing negotiated price levels, in which case spending based only on standardized prices were reported. In these cases, although information on actual spending incurred for health care services was not available, these adjusted TCC measures give a more accurate description of medical resource use. Because price variation is limited, a clearer picture of the volume and intensity of services delivered is left. [51]

One issue that needed to be dealt with when developing the price-adjusted measures was that each site was using a different standardized pricing list. In order to facilitate the ability to develop nationally consistent measures using standardized prices, a publically available standardized price list should be made available.

[51] Volume refers to the number of services rendered, whereas intensity refers to the amount of resources required for each individual service.

Building a More Robust Measure Set over Time

Moving forward, developing a more robust measure set over time will be essential for effective heath care evaluation. Claims measures are very useful—particularly for measuring utilization and total cost of care—but are limited in their clinical utility. As HIT adoption increases, community initiatives should gradually move toward clinically-enriched measures rather than relying on claims data alone so that the health outcomes of interventions can be assessed in addition to process-based measures of care quality. Exhibit 43 illustrates how the clinical utility of performance measures increase as measures move from claims-based process measures (e.g., receipt of HbA1c tests) to clinically-enriched outcome measures (e.g., HbA1c values) to patient-reported and EHR-sourced data (e.g., self-reported health status and daily blood glucose level fluctuations).

Exhibit 43: Developing a More Robust Measure Set Over Time as Health IT Adoption Increases

	Quality Measurement and Reporting	Cost Measurement and Report
Electronic and Personal Health Record Data	• A1c value + daily blood glucose deviations • Quality of life and functional outcomes for diabetes	
Administrative Data & Specific Clinical Data (e.g., Lab Values)	% of Type 2 diabetics with A1c <7.0%	
Administrative Data	# of A1c tests ordered	• Total costs • Hospital readmission for diabetes

At the same time, this project demonstrated that HIT adoption rates need not be a barrier to reform implementation. Meaningful and effective performance measurement is possible at every stage of HIT adoption (For a table of performance measures at every stage of HIT adoption, see Appendix.

The Reality of Policy Evaluation in a Post-ACA Environment

Unlike scientific experiments, which take place in carefully controlled environments and test for marginal effects of isolated interventions, this project demonstrated why large-scale community-based reform efforts do not easily lend themselves to pristine evaluations.

Community-Based Interventions Interacting with Multiple Initiatives

This project demonstrated how difficult it can be to isolate the effects of particular interventions. In Vermont, for example, reform efforts were already underway for improving the quality of care amongst diabetic patients prior to the implementation of the Blueprint program. Going forward, this will make it difficult to separate which observed improvements in diabetic care and outcomes were the result of the Blueprint project and what were due to the success of the programs already underway in Vermont. The likelihood that any particular intervention in isolation will be effective In order to achieve the magnitude of savings and quality improvements anticipated by the Affordable Care Act is low. For instance, preliminary results from the Physician Group Practice Demonstration suggest that the most successful models had multi-dimensional initiatives that were not just limited to a specific chronic condition or specific intervention. One of the most successful demonstration sites, the Marshfield Clinic,

exceeded quality measures, generated significant savings, and earned a performance payment from CMS by synthesizing chronic care management programs, performance feedback to providers, HIT, and physician/nurse regional teams.[52]

Such a multi-faceted approach presents new challenges from an evaluation standpoint because it may be difficult to isolate the effects of specific payment reforms. Because multiple payment reforms are being implemented and are likely to continue to be implemented in simultaneous and overlapping ways, dealing with this problem is an evaluation necessity and an area where further methods development would be very beneficial.

Identifying Comparison and Intervention Groups

This project has illustrated the challenges in identifying comparison populations and why using control counties and other geographically-based regions to evaluate reforms may not be a long-term, viable strategy to support the robust evaluation of delivery and payment reforms on a large-scale basis, for at least two reasons. First, the use of control counties does not permit program expansion, since pilots depend on control areas not participating in the pilot in order to document quality and cost effectiveness. Second, it is becoming increasingly difficult to identify "clean" equivalent comparison county populations that are not somehow involved in other quality-improvement or related activities with the proliferation of multiple delivery and reform strategies being experimented with in the field. Hence, concerns about contamination of comparison patients and the legitimacy of comparisons of trends between control and intervention counties have become ongoing problems and identifying clean comparison counties can often be a time-consuming process.

In Vermont, different approaches were used for the Medicare population and the commercial population. Dartmouth chose to use weighted counties to identify comparison patients, while Onpoint used patient-level matched cohorts. A concern around Onpoint's approach was that patient-level matching would result in the exclusion of too many members of an already-small intervention group due to the inability to find a suitable match in the comparison group. While this did not create insurmountable limitations for this project, it could be of concern for other communities conducting analyses with less data. Additionally, Onpoint's matched cohort approach caused delays in identifying the comparison group for Vermont, as multiple versions of control matching had to be run before agreeing on the best version of matched controls to use.

Over time, moving toward methodologies that can compare actual performance from predictions based on historical trends of the intervention-specific population may provide a more timely alternative approach. Observational studies using electronic health information can be ideal for tracking patients over time and observing how their care quality and health outcomes vary as a function of different policies, such as changes in formulary designs and payment models, thus enabling efficient pre- and post- experimental research designs.[53, 54] Leveraging HIT and the electronic data infrastructure that is developing with ARRA funding, such a pre- and post- evaluation strategy could be used in support of

[52] Theodore A. Praxel, MD, MMM, FACP, CMS PGP Marshfield Clinic Experience. Retrieved December 13, 2011, from http://uwphi.pophealth.wisc.edu/healthpolicy/ebhpp/events/20080429/praxel.pdf.

[53] Schneeweiss S, Walker AM, Glynn RJ, Maclure M, Dormuth C, Soumerai SB. (2002). Outcomes of reference pricing for angiotensin-converting-enzyme inhibitors. *New England Journal of Medicine* 346(11):822-9.

Schneeweiss S, Walker AM, Glynn RJ, Maclure M, Dormuth C, Soumerai SB (2002). Outcomes of reference pricing for angiotensin-converting-enzyme inhibitors. *New England Journal of Medicine* 346(11):822-9.
349(23): 2224-2232

rapidly evaluating models rather than the post-hoc evaluation of treatment sites relative to control counties.

Staggered Start-up Dates
Cleanly defining the pre-intervention period from the post-intervention can be difficult because health reform interventions are often rolled out in different communities over time. For example, a pilot site might serve as an initial site for intervention before a reform effort is applied system-wide or to an entire state. Staggered start-up dates can lead to problems in defining pre-intervention and post-intervention data. The start-up dates for the Vermont and PCIP interventions were staggered over time by provider site, complicating pre-post evaluations of site interventions. Additionally, with any intervention there is usually a ramp-up period to full implementation. Many interventions need time to achieve their full intended impacts.

The approaches to dealing with staggered start dates employed in this project can inform the evaluations of other community-based reform initiatives. In Vermont, the start-up date for the Medicare population was established as the calendar year in which the Blueprint intervention became operational at a particular location. For the Medicare population, such an approach was viable because enough Medicare data was available for the period prior to launch. However, data for the commercial population were more limited and were not available going as far back in time. For that reason, start-up dates in the commercial population were established on a rolling basis starting with the month in which the intervention was started for each practice site. In turn, the pre-intervention period was defined as all the months prior to that particular month, going as far back as possible given the years of data available, to maximize the amount of pre-intervention data available to conduct a baseline assessment. In New York City, despite the varying start dates of PCIP providers, the pre-intervention period was defined consistently across Medicare and Medicaid. All patients attributed to PCIP providers that became EHR-enabled in October 2007 or later were included in the intervention group. This solution was viable in New York City because this method excluded only a small number of providers (mostly community health centers) who adopted PCIP's EHR system prior to October 2007.

This project demonstrated that perhaps the best approach to staggered start-up dates is to be site-specific, depending on how the intervention start-dates are distributed and what impact that may have on data availability, sample sizes and statistical power.

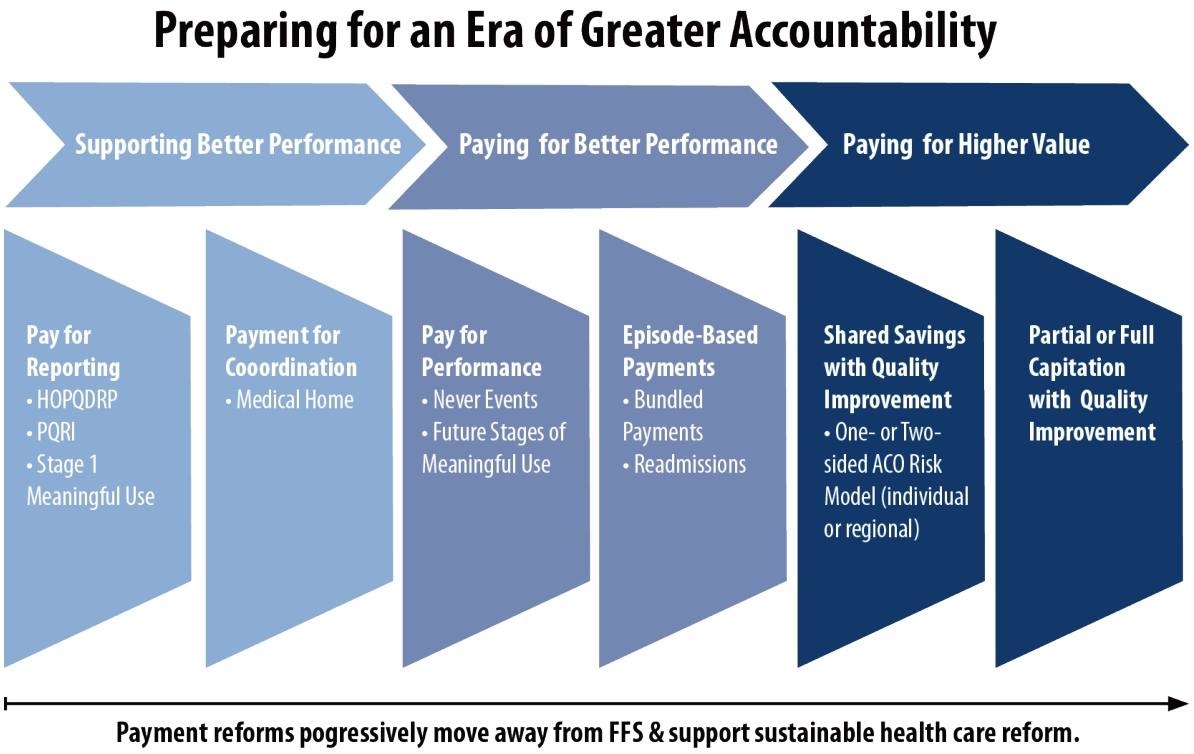
The Need for More Timely Medicare Data
Finally, the lag time in the release of Medicare data had a major impact on the ability to analyze the effects of the site interventions relative to baseline. The most recent Medicare data available at the start of this project was 2007 data, preventing this project from looking at both pre- and post-intervention data.

Given the movement toward payment models that increasingly hold providers accountable for the result of their care provision, the more timely release of Medicare data—not only for retrospective evaluation purposes but for more prospective care coordination, risk-stratification, and near-real-time performance monitoring to aid in continuous quality improvement purposes—will be critical.

Conclusion

All developed countries are grappling to control health care expenditures, yet the situation in the United States is especially acute. The 2011 Medicare Trustees report and the Congressional Budget Office's long-term fiscal outlook both make clear that growth in total health care expenditure in the United States remains unsustainable.[55] As health care costs continue to rise at an unsustainable rate, pressure on communities, and the country, to improve the quality of care being delivered while slowing cost growth will continue to intensify. In response to this challenge, communities have started to develop and implement initiatives to promote accountable care, often through the effective use of HIT. Broadly speaking, such reforms involve establishing, in the delivery system, greater accountability for cost and quality, through changes that include movement away from fee-for-service payments based on the volume and intensity of services regardless of their quality (see Exhibit 45), to other payment models that more clearly link financial rewards to demonstrably higher value care.

Exhibit 45. Preparing for an Era of Greater Accountability



The transition to accountability-based payment will not be easy. Communities that plan for this change will likely find the transition more orderly and effective. Planning around evaluation methods will be a critical piece of such reform in order to identify best practices and ensure that reform interventions are having their desired effects. This project was intended to facilitate reform efforts through the development of sound evaluation baselines against which future impacts can be assessed and can be used to identify effective community-based HIT-enabled interventions and encourage their wider distribution. While developing such an infrastructure is a resource-intensive endeavor, hopefully

[55] https://www.cms.gov/ReportsTrustFunds/downloads/tr2011.pdf

communities will be able to realize meaningful care improvements that also result in cost savings, a portion of which could then be re-invested back into the community to help finance such investments. Incremental steps like these are critical for initiating a virtuous cycle of quality improvement activities that can help finance further investments in the infrastructure needed to sustain them and are necessary steps along the way towards payment systems that ultimately support providers and communities in promoting high-value accountable care.

Federal investments in HIT have prompted discussions about how electronic health data can be used to support a range of delivery system reforms to improve the coordination and efficiency of care. In order to develop better evidence on how payment reforms and HIT can be used to improve care, a common template is needed to measure the effectiveness of HIT interventions across different payers and communities. The results of such evaluations can inform best practices and the development and implementation of a broader framework for improving the value of health care that is delivered, especially through the effective use of HIT. Extrapolating lessons learned from the evaluation of PCIP, Vermont, and WHIO, the template developed through this project and the challenges encountered along the way provide important insight for the evaluation of future community-based health care reform interventions.

Using a common set of measures to facilitate system-wide and cross-site evaluation of interventions is key to developing better evidence on how quality can be improved based on reform efforts. While community-led metrics can be an important piece of evaluation, a common template is needed to more rapidly evaluate the impact of community-led reforms across payers and thus help prepare communities for an era of greater accountability. A consistent way to measure total cost of care across payers is also critical in order to reliably assess the value of care patients receive and to monitor progress in bending the health care cost curve.

Variation in the health objectives and nature of the initiatives in PCIP, Vermont, and WHIO presented challenges in developing a common template to be used across all three communities. The diversity of these sites models the diversity of other community-based health improvement. The differing formats, elements, and quality of data from different payers make conducting analysis across payers challenging. These challenges suggest the need for a distributed approach to conducting analyses in electronic health care data, so individuals who are most familiar with the intricacies of particular data sources remain very close to the analysis.

The findings of this study demonstrate how important it is to look across cost, quality, and utilization metrics to assess the effectiveness of interventions. Looking at measures in isolation might lead to inaccurate assessments of the value of care that a community is providing. For example, while the cost of care might be lower in some areas, such lower cost might come at the expense of the quality of care being provided.

As the United States moves toward payment models that hold providers more accountable for the outcomes of the care they provide, policy makers will need to link patients to providers. Given the importance of multi-payer evaluations for reasons discussed earlier in this report, the ability to consistently identify providers across multiple payer data sources will also be required. Through this project, provider-matching and patient attribution methods have been developed that can help communities accomplish this.

This project led to the development of implementable measures of cost of care measure. In communities that have access to allowed charges, a measure that assesses the actual cost of care was developed. In communities in which those data are not generally available, total cost of care was monitored using standard prices. These total-cost-of-care measures, along with the suite of multi-payer utilization metrics developed for this project, can also be useful in assessing progress towards bending the Medicare cost curve. If CMS can use these measures consistently in their demonstration programs, it will allow for comparisons between demonstration sites and sites that are not participating in those demonstrations. Such evaluations will be helpful for benchmarking and facilitating system-wide improvements in health care value.

Meaningful and effective performance measurement is possible at every stage of HIT adoption. However, as communities continue to invest in an HIT infrastructure, evaluations should move toward a more robust measure set. While claims measures are useful for assessing utilization and total cost of care, they are limited in their clinical utility. As HIT adoption increases, community initiatives should gradually move towards clinically-enriched measures that provide a more complete and accurate window on clinical outcomes.

New evaluation approaches that account for the lack of randomization and control that are typical of community-based reform initiatives may be necessary to assess the impact of these interventions. In the meantime, baseline assessments that document the impact of previously implemented and ongoing reforms and identify site-specific areas to target in the next steps of reform will be important to guiding future reform initiatives. Baseline assessments, such as the ones conducted for this study, are an important step to enable policy makers to monitor and learn from early experience with health care reform, both so that reform efforts become more effective locally, and so that experience in one community can increase the effectiveness of efforts in others.

Acronyms

AF4Q	Aligning Forces For Quality
AHRQ	Agency for Healthcare Research and Quality
ASC	Ambulatory Care Sensitive Conditions
ASO	Administrative Services Only
ASR	Age, Sex, Race
BTE	Bridges to Excellence
CCT	Community Care Team
CDC	Centers for Disease Control and Prevention
CHF	Congestive Heart Failure
CMS	Centers for Medicare & Medicaid Services
COPD	Chronic Obstructive Pulmonary Disease
DOHMH	Department of Health and Mental Hygiene
DME	Durable Medical Equipment
DRGs	Diagnosis-Related Groups
E&M	Evaluation and Management
eCW	eClinicalWorks
ED	Emergency Department
EHR	Electronic Health Record
ERG	Episode Risk Group
ETG	Episode Treatment Groups
HCC	Hierarchical Condition Category
HEDIS	Healthcare Effectiveness Data and Information Set
HIE	Health Information Exchange
IP	Inpatient
IT	Information Technology
NCQA	National Committee for Quality Assurance
NPI	National Provider Identification
NYC	New York City
NYU	New York University
PCIP	Primary Care Information Project
PCP	Primary Care Provider
PMPM	Per-Patient-Per-Month
RVUs	Relative Value Units
TCC	Total Cost of Care
TCNY	Take Care New York
UPIN	Unique Provider Identification Number
VHCURES	Vermont Healthcare Claims Uniform Reporting and Evaluation System
VITL	Vermont Information Technology Leaders
WHIO	Wisconsin Health Information Organization

Acknowledgments

Project Team

The Engelberg Center for Health Care Reform at Brookings

Mark B. McClellan, MD, PhD, Director
S. Lawrence Kocot, JD, MPA, LLM, Deputy Director
Karen Matsuoka, PhD, Research Director
Cary Sennett, MD, PhD, Managing Director
Molly Chidester, Research Assistant

The authors thank Kristina Hanson Lowell, Mark Zezza, Marisa Morrison, and Rory Thompson for their contributions to this research project.

Special Thanks

The Dartmouth Institute

Dan Gottlieb, MS, Senior Analyst/Statistician
Jeremy Smith, MPH, Analyst

The New York Primary Care Information Project Team

Samantha De Leon, PhD, City Research Scientist for the Primary Care Information Project
Tod Mijanovich, PhD, Research Assistant Professor, New York University
Sarah Shih, MPH, Executive Director of Program Evaluation for the Primary Care Information Project,

The Vermont Blueprint for Health Team

Karl Finison, Director of Analytical Services, Onpoint Health Data
Craig Jones, Director of the Blueprint for Health Program, Vermont Department of Health

The Wisconsin Health Information Organization Team

Julie Bartels, CEO, Wisconsin Health Information Organization
Dan Dunn, PhD, Senior Vice President of Business Solutions, OptumInsight (Ingenix)

* * *

This project was funded through a generous grant from

The Markle Foundation