Unpacking Dimensions of Performance in Healthcare Delivery

Unpacking Dimensions of Performance in Healthcare Delivery

A dissertation presented

by

Philip Saynisch

to

The Committee on Higher Degrees in Health Policy

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Health Policy

Harvard University

Cambridge, Massachusetts

April 2019

Dissertation Advisor:

Professor Robert S. Huckman

Author:

Philip Saynisch

Unpacking Dimensions of Performance in Healthcare Delivery

Abstract

In complex healthcare settings, the optimal choice of treatment can be highly ambiguous. As a consequence, the marginal patient may face radically different care depending upon the choice of provider. Moreover, the complexity of treatment in these settings heightens information asymmetries between patient and provider at the same time as the stakes are at their highest. In light of these challenges, this dissertation aims to identify the drivers of performance for healthcare organizations and providers, and the mechanisms by which this performance can change over time. It does so in two contexts: the decision-making and surgical performance of kidney transplant teams, and in primary care practices adopting the patient-centered medical home model.

Chapter One (work with Robert S. Huckman and Nikoloas K. Trichakis) explores how kidney transplant center volume impacts two dimensions of performance: decision-making and surgical execution. Using learning curve models, I find evidence that larger transplant centers have better post-transplant outcomes. However, patients at these centers are less likely to receive a better organ and more likely to die or be removed from the transplant waitlist after an offer is declined, indicating lower-quality decision-making. This tension between improved execution and reduced decision quality implies that practice may not make perfect in complex medical decision-making.

In Chapter Two (work with Guy David and Aaron Smith-McLallen), I contribute to the literature on the impact of the patient-centered medical home (PCMH) model. Using data on the specific capabilities adopted by practices, I employ hierarchical clustering to group practices based on their approach to the

PCMH. By evaluating the clusters as separate interventions, I find that treating the PCMH as a single model obscures important variation in patient outcomes.

Chapter Three (work with Guy David, Aaron Smith-McLallen, Spencer Luster and Ravi Chawla) focuses on how the PCMH model effects one important patient outcome – medication adherence. To do so, I identify a subset of six PCMH features which were most directly related to that outcome. We find that adoption of the PCMH model was associated with improvements in medication adherence, and that these gains were concentrated in practices that adopted four or more of the adherence-related capabilities.

In Chapter Four (work with Guy David, Benjamin Ukert, Abiy Agiro, Sarah Hudson Scholle and Tyler Oberlander), I use the approach outlined in Chapter Two to study the PCMH model, but extend the work to a larger sample of patients and practices, with the sample covering more years and geographic regions. We find an overall effect on healthcare utilization (including an 8% reduction in total expenditures), but also identify significant heterogeneity by cluster, with a reduction in emergency department utilization driven entirely by one group of practices emphasizing the adoption of enhanced electronic communications capabilities.

Table of Contents

List of Tables

List of Figures

Acknowledgements

First and foremost, I want to thank my committee – Rob Huckman, Guy David and Ariel Stern – for their incredible support and encouragement, both during the dissertation writing process and throughout my academic career. The knowledge that I could rely on this incredible team of scholars not only for advice on methodology or theory but also in navigating the challenges of life as a graduate student was a tremendous comfort, and I am incredibly grateful for their contributions.

During the summer before I started the PhD program, Rob and I were paired to work on a pre-PhD research project. While intended primarily to introduce incoming students to the data and resources available at HBS, I immediately realized that not only did Rob and I have similar interests, but that he would be an incredible advisor as well. I have continued, without pause, to learn from him since. I began working as a research assistant for Guy at nineteen while I was an undergraduate at the University of Pennsylvania, and graduated to co-author and collaborator – but continue to view him as a mentor and friend. Ariel's contributions have been instrumental in my success as well, providing invaluable insights into how to present my work to an audience that crosses disciplinary and subject matter boundaries.

I also want to thank the Health Policy and HBS Doctoral Programs staff – Debbie Whitney, Colleen Yout, Marais Canali Young, Jen Mucciarone, Angela Valvis, LuAnn Langan and former staffers Karla Cohen, Ayres Heller and Jessica Penca Livingston. All of these people have been tremendously important in my success at Harvard, providing not only access to resources and guidance about rules and regulations, but also making me feel like I had a home (or, in this case, two homes) at an unfamiliar and sprawling institution.

Special thanks are due to Nikos Trichakis. While not a member of my committee, Nikos' contributions were essential to my understanding of the kidney transplant environment and fundamentally shaped my thinking on these important questions. I am indebted to Richard Zeckhauser and Peter Reese as well for sharing their insights with respect to institutional details and analytic considerations surrounding organ

transplantation. The team at UNOS also provided extensive support, and I am grateful for their assistance. The data used in the kidney transplant work were supplied by UNOS as the contractor for the Organ Procurement and Transplantation Network (OPTN). Per their disclosure requirements, the interpretation and reporting of these data are the responsibility of the authors – that is, me – and in no way should be seen as an official policy of or interpretation by the OPTN or the U.S. Government.

My work on the patient centered medical home has also benefitted from the contributions of a diverse and talented team. My co-authors Abiy Agiro, Ravi Chawla, Spencer Luster, Tyler Oberlander, Sarah Hudson Scholle, Aaron Smith-McLallen and Benjamin Ukert all deserve acknowledgement here for their hard work and invaluable insights. I also want to thank Somesh Nigam, Cindy Wang, and Candace Gunnarsson for data support and helpful suggestion in framing questions and analyses.

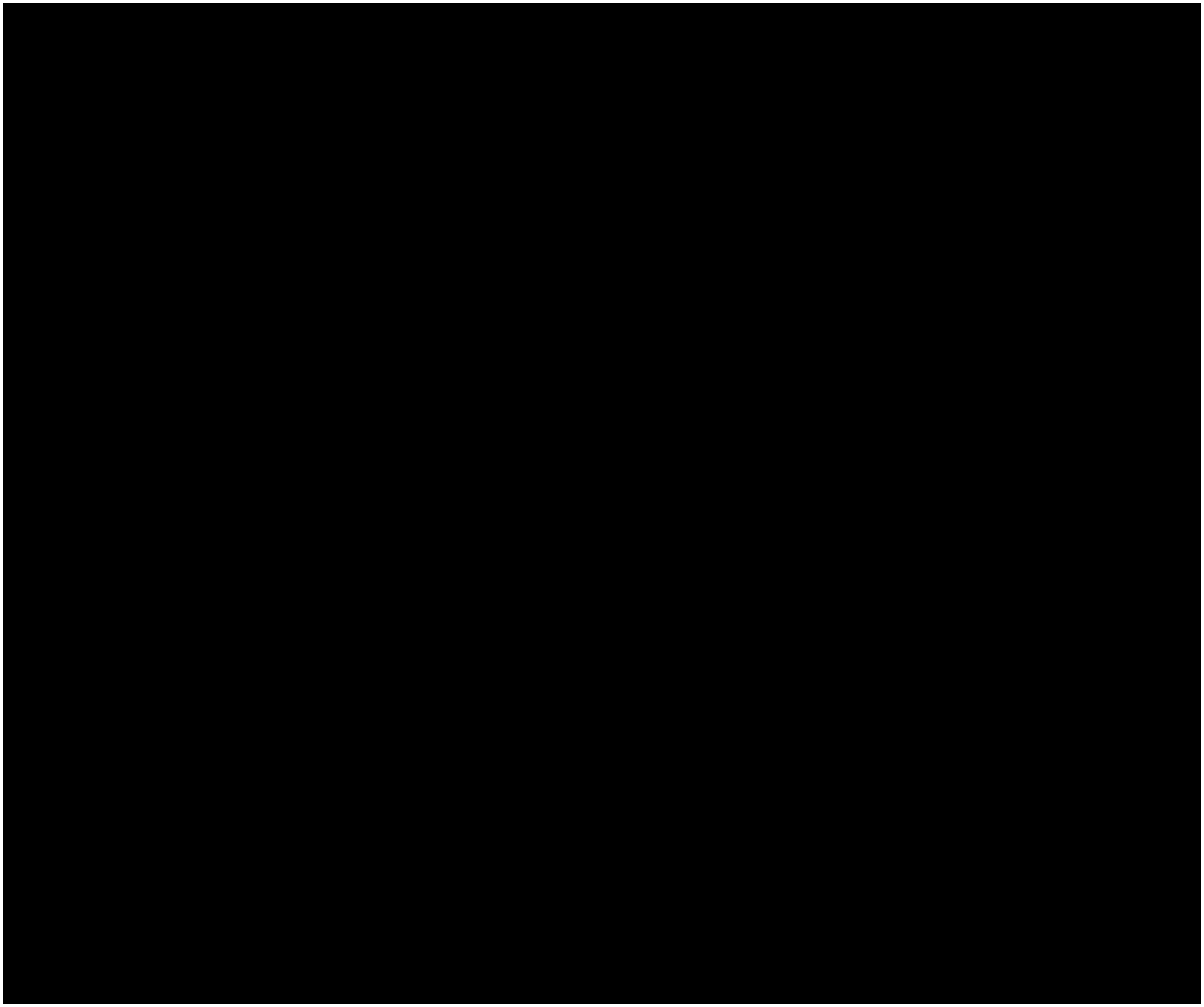
The work presented in this dissertation has also benefitted extensively from comments from colleagues at a number of academic conferences and seminars. I thank participants at the Production and Operations Management Society 2017 and 2019 Annual Meetings, the American Society of Health Economists Annual Meetings in 2016 and 2018 and the Harvard Health Policy seminar.

I also want to thank my parents, Drs. Michael and Dorothy Saynisch. There are no words to adequately repay them for the unconditional love and support I have received from them over the past 32 years. Having two parents who earned PhDs themselves – and whose experience I could look to for advice and encouragement – was an incredible comfort over the last six years. Finally, I want to thank Heather Robertson for being a brilliant, patient and kind partner, and for bringing Emma “The Dog” Robertson into my life.

Chapter 1

Decomposing Volume's Impact on Performance: Lessons from Kidney Transplantation

with Robert S. Huckman and Nikolaos K. Trichakis



Chapter 2

The Economics of Patient-Centered Care

with Guy David and Aaron Smith-McLallen

2.1 Introduction

The increasing prevalence of chronic illnesses and mounting associated costs are major concerns for the US healthcare system. The bulk of efforts to control health care costs and improve the quality of care have focused on two areas familiar to economists: regulation and alternative payment models. Efforts to restrain the growth of healthcare costs through regulation included licensure for health care providers and capital restrictions (e.g. state certificate of need laws). Similarly, alternatives to fee-for-service reimbursement, such as capitation, prospective payments, and bundled payments for episodes of care have also been advanced as potential solutions and implemented with various degrees of success. In addition to these reforms, a third approach has focused on reorganizing healthcare organizations to increase their population health and chronic illness management capabilities, reduce provision of low value care, improve care coordination, and raise patient and provider satisfaction.

One such model for reorganizing healthcare organizations which has generated considerable attention is the Patient-Centered Medical Home (PCMH). This reform consists of a suite of primary care improvements including assignment of patients to a personal physician responsible for directing "whole person" care, adoption and use of health information technology and expanded patient access to care providers (American Academy of Family Physicians, American Academy of Pediatrics, American College of Physicians, & American Osteopathic Association, 2007).

Despite numerous pilot projects and extensive efforts at evaluating the PCMH, the evidence regarding the model's impacts on patient experience, utilization and expenditures remains mixed. This study aims to

improve the present understanding of how the PCMH model may increase the reliance on primary care and reduce downstream utilization of specialist care as well as emergency department visits and hospitalizations. It does so through the use of a unique dataset containing detailed data on specific capabilities for medical home practices, which has never previously been linked to patient-level claims data. Because primary care practices can achieve PCMH recognition from the National Committee for Quality Assurance (NCQA) by adopting a self-selected subset of practice improvements, a binary categorization of practices as medical homes or not could obscure substantial variation in implementation. One recent study described substantial variation in how practices implemented the PCMH model, and its authors point out that further research is needed to know whether these different approaches have varying impacts on patient outcomes (Tirodkar et al., 2014). This echoes a more general call in economics to engage in "mechanism experiments," studying not only whole policy interventions but also attempting to identify the specific channels by which these programs yield improvements (Ludwig, Kling, & Mullainathan, 2011). By documenting which specific PCMH capabilities were present in the recognized practices, we intend to assess whether these capabilities have differential impact on patient interaction with the healthcare system, potentially clarifying some of the conflicting results from prior studies.

This paper uses data on 152,093 patients over six years (370,764 patient-years in total) covered by a single large, private insurer in southeastern Pennsylvania. These patients were treated in 104 practices which gained recognition as medical homes between 2008 and 2012, with healthcare utilization and expenditures tracked through 2013. The dataset describing the specific PCMH components which practices had in place at the time of initial recognition is extremely detailed, including scoring based on 127 individual "factors" of implementation.[20] As a consequence, the number of dimensions of interest exceeds the number of practices with which to study them. Moreover, the specific functional form of the primary care production

[20] The 127 individual NCQA factors span 139 binary components.

function is not known, and there may be important interactions between PCMH factors in determining patient outcomes.

To address these issues, we group medical home practices into one of three types using a hierarchical clustering approach. The algorithm used here starts with *N* clusters of 1 practice, and sequentially groups practices based on their similarity in terms of implemented PCMH "factors." This dramatically reduces the dimensionality of the problem described above, as we are able to include only indicators for cluster identifiers in the regression analyses. In addition to making the analysis possible, this approach substantially reduces the risk of overfitting, as the clusters are defined without reference to any of the outcomes of interest. Using this approach, we find substantial heterogeneity in implementation across clusters, with different areas of focus (generally summarized as the "basic model," which passes only the minimum PCMH requirements; a "patient-facing" emphasis, focusing on enhanced population health management; and a "physician-facing" variant, stressing enhanced access, decision support and data reporting). Moreover, we find that both analyses which treat the PCMH model as an undifferentiated intervention and alternative specifications that use the level of PCMH recognition to differentiate between practices miss significant variation in effects on patient expenditure and utilization outcomes, which become apparent when the PCMH model is instead evaluated in terms of the performance of clusters of similar practices. Additionally, we find that this pattern – heterogeneity in PCMH implementation and subsequent differential effects on patient outcomes – is not explained away by practice-type mix or prevalence of chronic illness across practices.

The rest of the paper is organized as follows: the following section describes the patient centered medical home model and discusses previous efforts at evaluating its effects on patient outcomes. Section 2.3 discusses the sources of patient and primary care practice data used in this study. Section 2.4 describes our empirical approach, covering both the clustering technique and regression analysis used to assess patient-level data. In Section 2.5, we present our results. We find that practices achieve PCMH recognition with different combinations of eligible features, and that this variation is not captured by previously-used

summary measures such as PCMH level. Additionally, when considered as a single intervention, PCMH adoption explains little of the variation in patient outcomes; however, when practices are analyzed as belonging to a commonly implemented subtype, a relationship between PCMH adoption and patient outcomes emerges. Section 6 includes discussion of these results, and Section 7 concludes.

2.2 The Patient Centered Medical Home

An extensive body of literature points to a number of problems with the status quo approach to organizing primary care - for example, care is often structured to address acute health issues, rather than to manage on-going concerns (Bodenheimer, Wagner, & Grumbach, 2002; Wagner, Austin, & Von Korff, 1996). Specialist and procedural services such as diagnostic imaging are reimbursed at higher rates compared to core primary care activities aimed at disease management, and physicians may not be paid at all for work to coordinate care outside of primary care practices (Bodenheimer & Pham, 2010). Additionally, fewer than half of office-based physicians had an electronic health record (EHR) system in place as of 2008, when the PCMH initiative considered in this study began (Hsiao & Hing, 2014), potentially leading to difficulty monitoring the care and condition of patients with chronic illnesses.

The patient centered medical home is the leading model currently being advanced to address these significant problems in primary care. The medical home provides "whole person" care, aimed at treating acute needs as well as focusing on broader goals like population health management, coordination of care across sites, improved patient engagement, implementation of evidence-based care, using health information technology, expanded practice hours and improved patient-provider communication.[21]

These goals were operationalized by the National Committee for Quality Assurance (NCQA), an organization which has produced a recognition checklist for practices seeking accreditation as patient centered medical homes. In recent years, these guidelines have become the standard recognition criteria

[21] The origins and evolution of the PCMH model have been extensively documented by health services researchers (Friedberg et al., 2009; Kilo & Wasson, 2010).

used in evaluations of the PCMH model (Cassidy, 2010; Friedberg, Lai, Hussey, & Schneider, 2009). By October 2014, 11,058 sites of care and 55,156 clinicians were recognized by NCQA as PCMH (roughly a quarter of all practicing primary care physicians in the U.S.). The 2008 NCQA guidelines, which were used to evaluate the practices included in this study, provide 127 specific action items (or "factors,") which are grouped into 30 "elements," which are further grouped into nine "standards;" practices accrue points based on the number and type of PCMH factors which have been implemented, with different weightings depending on the specific area of practice improvement. Practices achieve PCMH recognition at one of three levels by satisfying two sets of requirements: first, they must receive at least 50% of the possible points in five or more of ten "must-pass" elements; second, they must receive 25 or more total points (out of a possible score of 100). Practices qualify for Level 1 recognition with 5 or more must-pass elements and 25 points; Level 2 is reached when practices pass all 10 must-pass elements and have accrued at least 50 points; and Level 3 is reserved for practices satisfying all 10 must-pass requirements and receiving a total score of 75 points or more (National Committee for Quality Assurance, 2008). An overview of these standards and elements appears in Appendix Table B1.

The PCMH model is expected to lead to better care management and subsequently to a reduction in utilization of high-cost, high-intensity care such as hospitalizations or emergency department (ED) visits (Hearld & Alexander, 2012). By one estimate, as many as 27% of ED visits could have been effectively treated in an office-based setting (Weinick, Burns, & Mehrotra, 2010). The evidence on the PCMH model's impact on patient outcomes to date has been mixed. Several studies of the PCMH model have found evidence of reduced total expenditures (DeVries et al. 2012; Paustian et al. 2013) and lower utilization of high-cost medical services such as hospital admission and ED visits (Reid et al., 2010; David et al., 2015; Rosenthal et al., 2015, 2016). Others (including systematic literature reviews) found the PCMH to have limited or no effect on quality, utilization or expenditures (Friedberg et al., 2014; Peikes, Zutshi, Genevro, Parchman, & Meyers, 2012; Jackson et al., 2013; Peikes et al., 2011). All these studies analyze

implementation of the PCMH model as a whole, rather than studying the specific components practices adopted.

2.3 *Data*

2.3.1 Program Description

This study analyzes data from HMO enrollees in a single large, commercial insurer in Pennsylvania, where an official Chronic Care Commission partnered with private payers to promote the PCMH model. This insurer actively encouraged primary care practices to seek PCMH recognition (level 1 or higher), offering support services as well as financial compensation to practices which implemented the PCMH model. Early adopters received payments in order to defray the costs of PCMH infrastructure In addition, level 3 PCMH practices receive an increase in reimbursement of $3.00 per-member per-month (PMPM), while level 2 and level 1 PCMH practices receive $2.00 and $1.25 PMPM, respectively.[22] The insurer offered a number of clinical support services as well: practices gained access to the American College of Physicians' Practice Advisor tool, which provides guidance on care, workflow and practice organization, as well as a PCMH resource library (provided by NaviNet), and a variety of clinical reports.

The use of HMO enrollees was driven by the need to assign patients to primary care providers in a way that is not reliant on their care utilization pattern. In contrast, Preferred Provider Organization (PPO) enrollees covered by this insurer are attributed to physicians based on their encounters, which serve as an outcome of interest in our study. Moreover, using attributable PPO members excludes members with little or no interaction with healthcare providers, a primary target population of the PCMH model.

Nonetheless, use of only HMO enrollees may limit the generalizability of this work, and it is difficult to assert if our results would be stronger or weaker for PPO members, as the effects of plan type on the PCMH effect are ambiguous. In an HMO, there is already substantial gatekeeping, and patients may already be

[22] Note that this boost is large, when considering that practices receive an average of $16.14 (PMPM) for commercially insured patients and an average of $25.89 (PMPM) for Medicare Advantage patients.

steered toward the most appropriate care, which would reduce the size of the estimated effect. On the other hand, in a PPO plan lacking utilization review, patients can often see specialists without prior authorization, meaning that physicians have limited ability to steer patients.

2.3.2 Recognition Data

In order to achieve recognition, practices used an online platform to submit documentation on implementation of PCMH capabilities for review. A small subset (5%) of practices went through a subsequence audit process (Tirodkar et al., 2014). Though data on practice PCMH recognition level had previously been available to researchers, the detailed data on specific practice capabilities had not. As a result, evaluations of the PCMH model have generally proceeded without detailed data on which capabilities practices had implemented. One exception is Friedberg et al. (2014), which used surveys of practices to assess which PCMH components were present. These survey results provided useful description of a small number of early PCMH implementation, but were not used to assess whether and how heterogeneity in implementation may have impacted patient outcomes. Another recent study (Tirodkar et al., 2014) used the detailed recognition data for descriptive purposes, identifying significant variation in PCMH implementation, even among practices recognized at Level 3. However, this study did not link the recognition scoring to claims data or any other information on patient outcomes, and states that further research to identify the highest-impact PCMH components is needed. For the following analyses, practice-level recognition data were obtained from NCQA. In order to facilitate this access, we have identified all the practices which adopted the PCMH model by September 2012. We contacted the practices to obtain consent for NCQA to release the detailed recognition data to us and followed up on all contacted practices intensively, making repeated attempts to contact non-respondents. The resulting dataset includes 139 binary dimensions of PCMH recognition for each practice, representing the most granular information used in the 2008 PCMH recognition process.

2.3.3 Selection Criteria

All practices in our analysis achieved PCMH recognition but differ in the timing of accreditation. Practices receiving recognition as a PCMH after September 2012 were excluded from consideration as this information did not exist at the time we began reaching out to practices to obtain consent for data release. Of 280 eligible practices[23], we received responses from 134 practices. We excluded 27 pediatric practices, since children have vastly different health needs and a different profile of health care utilization compared to the adult population. In addition, we excluded 3 practices that did not span the full six-year study period, leaving 104 practices whose data are included in this study. Additionally, during the study period, we identified 1,549 practices in our insurer sample, which never achieved PCMH recognition. In Table 2.1, we present summary statistics at the practice level, comparing three groups of practices: the 104 adult practices analyzed in this sample; the 118 adult practices which did not respond to the request for data; and the 1,549 adult "never adopter" practices.

The "never adopter" practices differ significantly from the responders and non-responding practices on nearly every dimension listed: they treat fewer patients in our sample, their patients are older and sicker, and they have higher health care utilization. Therefore, the never-adopters may not represent a comparable control group to the practices eventually achieving PCMH recognition. On the other hand, the responders and non-responders are much more similar, with significant differences in the share with coronary artery disease, the proportion with any ED visit, and in average professional expenditures. This suggests that selection into responding to the request for data release is likely not a major driver of our results.

Figure 2.1 represents the cumulative number of PCMH-recognized practices over time. The top line (short dashes) tracks the timing of recognition for all 280 PCMH in this market between March 2008 and September 2012. The dotted line below it tracks the 225 adult PCMHs (excluding 55 pediatric practices).

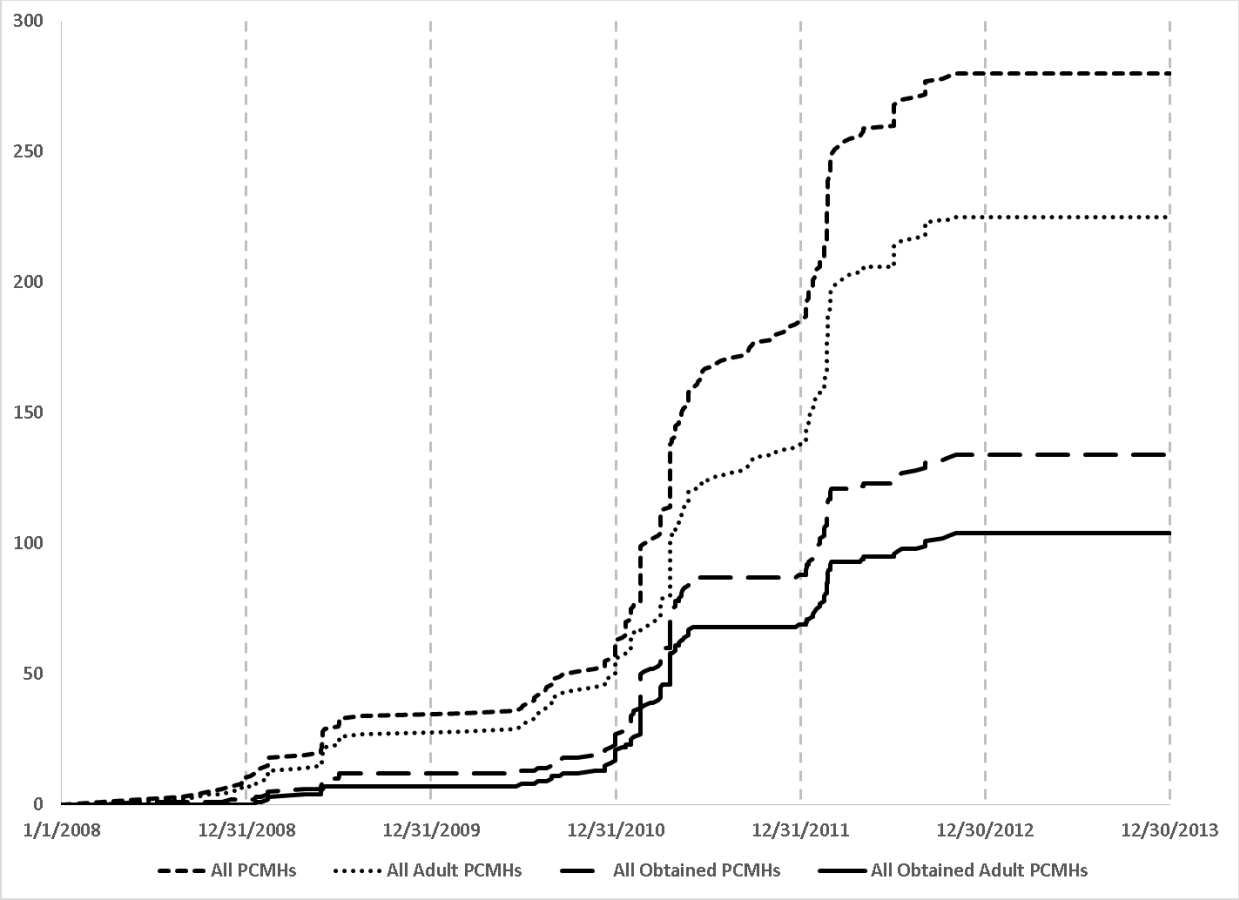
[23] All practices analyzed in this study were recognized under the 2008 PCMH guidelines. Though the 2011 recognition guidelines became available in March of that year, many practices used the 2008 standards through 2012 when our study period ends.

Table 2.1: Summary Statistics: Respondents, Non-Respondents and Never-Adopters

	Respondents	Non-respondents		Never-adopters
Practice Characteristics				
Number of practices	104	118		1,549
Internal Medicine	43	42		604
Family Practice	59	72		934
Pediatrics	0	0		0
Other	2	4		11
Average num members	763.91	759.49		248.34
	(671.58)	(778.08)		(327.68)
Total NCQA Points	74.07			
	(15.89)			
Patient Characteristics				
Average Age	45.35	45.25		46.46
	(6.81)	(7.27)		(9.54)
Average Risk Score	3.04	2.98		3.25
	(1.20)	(1.20)		(2.57)
Share Female	0.55	0.55		0.53
	(0.06)	(0.07)		(0.14)
Share Asthma	0.06	0.07		0.06
	(0.02)	(0.04)		(0.07)
Share Coronary Artery	0.05	0.06	**	0.07
Disease	(0.03)	(0.04)		(0.09)
Share Congestive Heart	0.02	0.02		0.03
Failure	(0.01)	(0.01)		(0.06)
Share Chronic Obstructive	0.03	0.03		0.04
Lung Disease	(0.02)	(0.02)		(0.06)
Share Diabetes	0.09	0.09		0.10
	(0.04)	(0.06)		(0.10)
Utilization				
Share Hospitalizations	0.08	0.08		0.10
	(0.03)	(0.03)		(0.10)
Share ED Visit	0.13	0.15	***	0.17
	(0.03)	(0.06)		(0.11)
Share PCP Visit	0.72	0.73		0.81
	(0.09)	(0.11)		(0.20)
Share Specialist Visit	0.39	0.40		0.43
	(0.09)	(0.11)		(0.18)
Average Professional	1779.64	1873.34	*	2087.95
Expenditures	(386.63)	(498.66)		(842.79)

+ p<0.1 * p<0.05 ** p<0.01 *** p<0.001

Figure 2.1: Cumulative Number of PCMH Practices by Day and Sample



The line below that (long dashes) tracks the 134 practices for which detailed NCQA recognition data was obtained, and finally, the bottom solid line tracks the final sample used in this paper. PCMH recognitions have accelerated over time with 4 recognitions in 2008, 7 in 2009, 10 in 2010, 46 in 2011 and 36 over 9 months in 2012. In 2013 all 104 practices had PCMH status. The recognition dynamics are similar across the four samples.

Patients enrolled in an insurance plan which did not require them to select a primary care physician (and who therefore could not be reliably attributed to any one practice) were excluded from the study. All patient-year observations include members with 12 months of continuous enrollment in the plan. Finally, to account for short-term disruptions around PCMH adoption and because all outcomes are observed at the patient-

year-level (e.g. expenditures per year), the year of PCMH recognition was excluded from this analysis.[24] Applying these exclusion criteria generated two datasets: the first includes 370,764 patient-year observations; and a second, referred to as "stable patient panel sample," is comprised of 130,923 patient-years for members enrolled in the same practice during all 6 years of the study period. The stable panel excludes potential selective entry and attrition of patients before and after PCMH implementation. Additionally, we identify patients with any of five chronic illnesses: congestive heart failure, chronic obstructive lung disease, coronary artery disease, asthma and diabetes.

2.3.4 Patient and Practice Covariates

A number of variables were used to control for patient attributes. In addition to demographic information like gender and age, each patient-year observation includes a risk score estimated using the Verisk Health DxCG Risk Solutions model, which incorporates clinical and demographic data for each patient (Verisk Health Inc., 2010). A number of practice descriptors, such as medical specialty were included. Summary statistics describing this pool of eligible patient-years appear in Table 2.2.

Several features of Table 2.2 warrant further discussion. First, note that patients in the stable "panel" sample (those with 6 years of continuous enrollment in the data) are somewhat older than in the "full" sample. This is due to the fact that the most common reasons for plan enrollment discontinuity are job switching, relocations, and enrollment in higher education programs, all of which happen more frequently at younger ages. Patients in the stable panel have comparable utilization to the full sample, with similar hospitalization rates, slightly higher use of primary and specialist care, and slightly lower ED use and expenditures. Patients with chronic illnesses are – as expected - older and have much higher rates of utilization and expenditures for professional services.

[24] Models including switch year data and controlling for both the year of PCMH implementation as well as the fraction of that year spent in PCMH status yield similar results, but are not as conservative due to potential issues taking place during the implementation year.

Table 2.2: Summary Statistics: Patient and Practice Characteristics

	All Adult Patients	Chronic Patients	Adult Panel	Chronic Panel	All Adult Patients			Adult Panel		
					Cluster 1	Cluster 2	Cluster 3	Cluster I	Cluster II	Cluster III
Observations	370,764	75,320	130,923	30,055	232,699	82,231	55,834	90,663	20,748	19,512
PATIENT CHARACTERISTICS										
Age	48.45	61.51	52.70	64.44	47.86	49.50	49.36	52.64	52.15	53.52
	(19.26)	(18.00)	(19.09)	(16.93)	(19.09)	(19.71)	(19.19)	(19.19)	(19.28)	(18.34)
Gender (Female)	0.56	0.54	0.57	0.54	0.57	0.55	0.57	0.57	0.55	0.58
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.49)	(0.50)	(0.50)	(0.49)
Risk Score	2.76	6.92	2.83	6.38	2.62	2.93	3.11	2.79	2.78	3.07
	(7.35)	(12.89)	(6.85)	(11.37)	(7.01)	(7.68)	(8.19)	(6.73)	(6.76)	(7.45)
Congestive Heart	0.02	0.10	0.02	0.09	0.02	0.02	0.02	0.02	0.02	0.02
Failure	(0.14)	(0.30)	(0.14)	(0.29)	(0.13)	(0.15)	(0.15)	(0.14)	(0.14)	(0.15)
Chronic Obstructive	0.03	0.15	0.03	0.14	0.03	0.03	0.04	0.03	0.03	0.04
Lung Disease	(0.17)	(0.35)	(0.18)	(0.35)	(0.16)	(0.18)	(0.20)	(0.17)	(0.17)	(0.20)
Coronary Artery	0.06	0.30	0.07	0.33	0.06	0.07	0.07	0.07	0.08	0.08
Disease	(0.24)	(0.46)	(0.26)	(0.47)	(0.23)	(0.26)	(0.25)	(0.26)	(0.27)	(0.27)
Asthma	0.07	0.33	0.07	0.29	0.06	0.07	0.08	0.06	0.06	0.08
	(0.25)	(0.47)	(0.25)	(0.46)	(0.24)	(0.25)	(0.27)	(0.25)	(0.24)	(0.28)
Diabetes	0.10	0.48	0.11	0.49	0.09	0.10	0.11	0.11	0.11	0.13
	(0.30)	(0.50)	(0.32)	(0.50)	(0.29)	(0.30)	(0.32)	(0.31)	(0.32)	(0.33)
UTILIZATION										
Hospitalization	0.08	0.20	0.08	0.18	0.08	0.09	0.10	0.08	0.08	0.09
	(0.28)	(0.40)	(0.27)	(0.39)	(0.27)	(0.29)	(0.29)	(0.27)	(0.27)	(0.29)
PCP Visits	0.81	0.94	0.83	0.94	0.83	0.77	0.81	0.83	0.83	0.83
	(0.39)	(0.24)	(0.38)	(0.23)	(0.38)	(0.42)	(0.39)	(0.38)	(0.38)	(0.38)
Specialist Visits	0.47	0.75	0.50	0.76	0.46	0.47	0.49	0.50	0.49	0.52
	(0.50)	(0.43)	(0.50)	(0.43)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)
ED Visits	0.13	0.19	0.12	0.17	0.13	0.14	0.14	0.12	0.12	0.13
	(0.34)	(0.39)	(0.33)	(0.37)	(0.34)	(0.34)	(0.35)	(0.33)	(0.32)	(0.34)
Professional	2,013.82	3,603.20	1,976.61	3,320.21	2,009.44	1,986.56	2,059.99	1,954.28	2,012.40	2,039.88
Expenditures	(6,691.28)	(7,190.32)	(4,848.64)	(6,158.84)	(7,095.74)	(5,479.74)	(5,841.15)	(4,603.22)	(5,238.89)	(5,464.21)
PRACTICE CHARACTE										
Number of Practices	104				50	33	21			
Total NCQA Points	74.07				82.00	55.90	83.73			
	(15.89)				(8.12)	(14.54)	(0.83)			
Internal Medicine	0.413				0.440	0.394	0.381			
	(0.495)				(0.501)	(0.496)	(0.498)			
Family Practice	0.567				0.560	0.545	0.619			
	(0.498)				(0.501)	(0.506)	(0.498)			
Pediatrics and Other*	0.019					0.061				
	(0.138)					(0.242)				

* "Other" practice type refers to Certified Registered Nurse Practitioner (CRNP) practices

2.3.5 Patient Outcomes Data

A summary table of patient outcomes, overall and by cluster, appears in the bottom panel of Table 2.2. Two sets of patient outcomes data are available for analysis in this study. First, patient healthcare utilization is measured in terms of hospital admissions, primary care physician (PCP) visits, specialist physician visits, and emergency department visits. We analyze data on the impact of the PCMH clusters on these utilization outcomes with respect to both extensive and intensive margins (that is, the probability of any encounter and the number of encounters conditional on having at least one, respectively). Additionally, per patient spending on professional services (that is, primary and specialist care) is considered in terms of the total amount of their medical claims that were considered eligible for payment.[25] Expenditures are available as a per-patient per-year total and are conditional on having non-zero expenditures in that year (reducing the sample from 370,764 to 322,539 patient-year observations).

2.4 Methods

2.4.1 Clustering Procedure

The evaluation of a practice-level intervention places our effective sample size (104 for which recognition data is available and which meet our inclusion criteria) below the number of PCMH recognition dimensions. This precludes the more straightforward analysis of the impact of individual facets of the PMCH model on patient outcomes. Therefore, the approach we take is to identify "clusters" of practices which implemented similar mixes of PCMH components, as measured at the "factor" level (the most granular level of data available on practice behavior). Clustering techniques are used in a variety of fields to develop taxonomies, including marketing (Punj & Stewart, 1983), strategic management (Ketchen Jr. & Shook, 1996) and health services research (Shortell, Wu, Lewis, Colla, & Fisher, 2014). The 2008 NCQA recognition process

[25] This amount is based on the insurance company's reimbursement rates, gross of coordination of benefits or subscriber liability.

includes 139 unique components.[26] While 2^{139} combinations of PCMH factors are technically possible, relatively few of these combinations are realized. Economies of scope are likely responsible for this fact.[27]

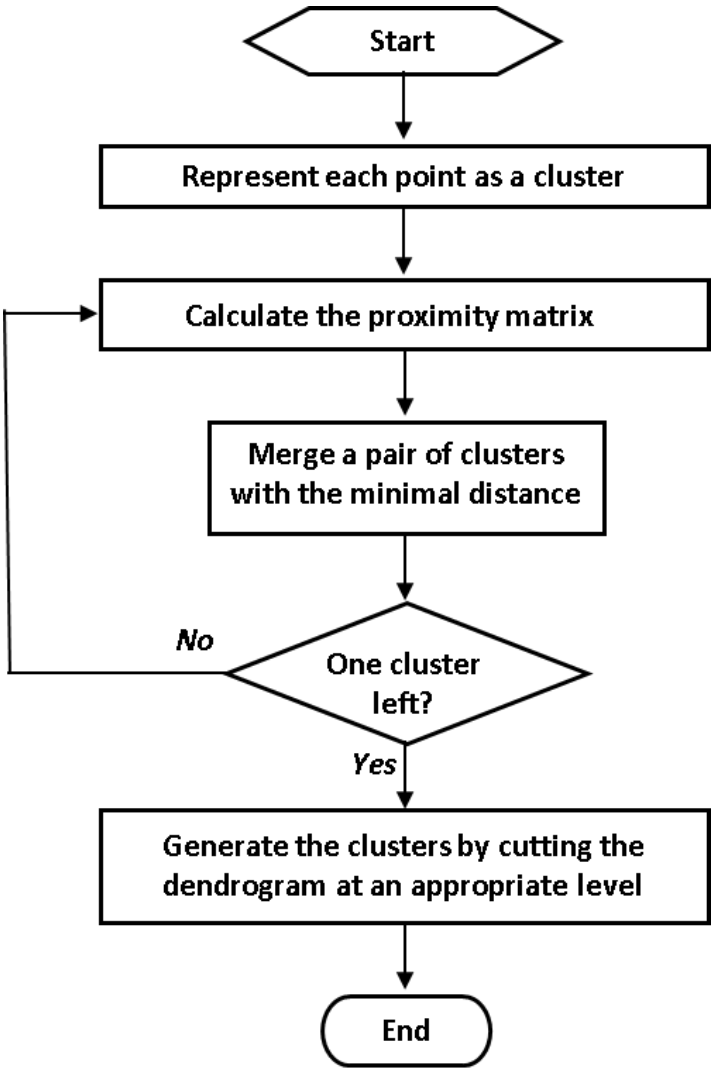
We employed an agglomerative hierarchical clustering approach with Ward's linkage, suited for clustering in high-dimensional space, to group similar practices (Sun, Wang, & Fang, 2012). Following Xu and Wunsch (2009) and Aggarwal, and Reddy (2013), this approach, summarized in the flowchart in Figure 2.2, begins with *N* clusters of size one, and sequentially combines the most similar practices into larger clusters. Every iteration involves calculating a proximity matrix, derived from a distance function, for the *N* clusters. The grouping process involves selecting in every iteration two subsets which, when united, generate the smallest increase in the value of the objective function. Given $k \in N$ subsets, this method allows for reducing *k* into *k*-1 mutually exclusive subsets by considering the union of all possible *k*(*k*-1)/2 pairs. This process is repeated until all subsets belong to a single group (Ward 1963). This iterative process can then be expressed in the form of a dendrogram, a branching diagram that visually represents the formation of clusters from maximally-similar practices.

Similarity of practices was measured using the Jaccard distance, and practices were grouped using Ward's method, which minimizes within-cluster variance. The objective function for the clustering process is given by: $E = \sum_{k=1}^{K} \sum_{x_j \in C_k} \left\| x_j - m_k \right\|^2$, where *K* represents the number of remaining subsets (i.e. clusters) and m_k is the centroid of cluster C_k, defined as $m_k = \frac{1}{n_k} \sum_{x \in C_k} x$, where n_k is the number of data points belonging to cluster *k* and *x* represents the value of a data point. The starting value of *E* is always zero, as

[26] Of the 127 factors, 117 were already in binary form (Yes/No) and ten expressed as either a percentage or offered four qualitative categories. Factors expressed as percentage were converted to a binary by assigning a threshold. For factors with heaping at an extreme value (0 or 100), we coded practices as achieving full implementation or not. For factors with a smoother distribution, a median split was employed. Qualitative factors were converted to a series of four dummy variables. In total, the clustering analysis was based on a total of 139 factors and sub-factors.

[27] For example, the individual factors under Standard 2, Element A ("basic system for managing patient data") list different pieces of information to be stored in an EHR, the majority of which would be present if any were. Other examples include "scheduling each patient with a personal physician for continuity of care" (Element A1, Factor 1) and "electronic system to order imaging tests" (Element B6, Factor 2).

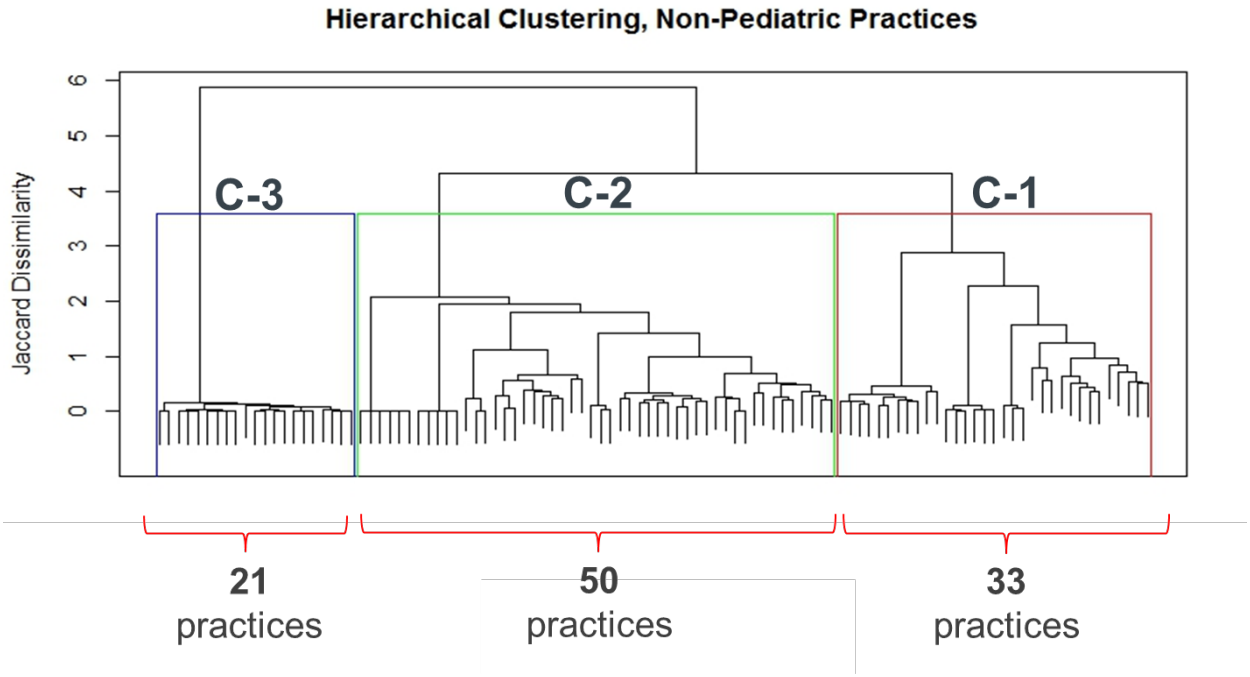
Figure 2.2: Hierarchical Clustering Algorithm Flow Diagram



n=1 (that is, each cluster contains a single data point and hence the cluster is equal to its centroid). At the end of the process *E* will equal some value that represents the sum of squared errors for all data points. Note that both the starting and end points are not informative, and only serve as a window with which the optimal path of grouping data-points into clusters can be obtained. Put differently, for each combination of two clusters (or subsets) *i* and *j*, we can calculate the corresponding change in *E*, the value of the objective function, as follows: $\Delta E_{ij} = \frac{n_i \cdot n_j}{n_i + n_j} \left\| m_i - m_j \right\|^2$. The smallest value of ΔE_{ij} is selected and the process is

then repeated for the remaining *K*-1 subsets, until a single cluster is reached. For the analysis which follows, a clustering solution was chosen through inspection of a dendrogram, which provides a visual representation of the increase in within-cluster distance after each aggregation. The root node of the dendrogram represents the entire data set and the intermediate nodes represent the extent dissimilarity (distance) at which the points were merged. Ultimately, a three-cluster solution was chosen.[28] The two most dissimilar clusters are Cluster 3 vs. Cluster 1 + 2, with a dissimilarity score of about 6, the next cutoff separates Cluster 1 and 2 with a dissimilarity score of above 4. The three-cluster solution is appealing as it contrasts with the three-levels of recognition for PCMHs and provides a competing partition of the practice space. Figure 2.3 shows a plot of the dendrogram used in this process.

Figure 2.3: Dendrogram for Hierarchical Clustering Approach



[28] The dendrogram in Figure 2.3 provides roughly equal support for a three- or four-cluster solution. However, in the four cluster solution, every practice in one of the clusters became PCMH recognized in the same year; a resulting collinearity between the treatment indicator and year fixed effects precluded use of the four-cluster solution.

This procedure defines the three clusters used in the following analyses, but does not provide any indication of how the practices in each cluster differ. In order to identify the distinguishing characteristics of each cluster, we use a one-way ANOVA to test whether mean scores for each element (that is, the level of aggregation above factors) differ by cluster.

The three-cluster solution was chosen in part to serve as a parallel to the NCQA-defined three-level typology of PCMH recognition, which is currently used to set PMPM reimbursement levels for participating practices. If the various components of the PCMH model were equally important to improving patient care, one would expect practices recognized at higher levels to strictly dominate those with lower levels of recognition, since greater numbers of total points and higher performance on must-pass elements is required to achieve each successive level. Moreover, there is good reason to expect that the clustering algorithm above will result in a partition that mimics these recognition levels. That is, if there is little within-level variation in the choice of recognition attributes, clustering should return the NCQA level classification. However, a high degree of within-level variation in implementation is likely to lead to a partition that crosses recognition levels and results in clusters that cannot be explained by NCQA recognition levels.

Varian (2014) outlines strategies for analyzing data when the number of regressors exceeds the sample size, including LASSO and spike-and-slab variable selection techniques. These approaches could be viewed as alternatives to the clustering analysis described above: rather than grouping practices into clusters with similar implementation, one could instead estimate models to identify the subset of PCMH factors which have the greatest impact on patient outcomes. However, these alternative approaches are best tailored to situations in which covariates are being chosen in order to maximize a model's predictive power with respect to a single outcome of interest, whereas this analysis considers a large number of patient outcomes. It would be possible to estimate separate models for each, but each LASSO (or similar) regression could identify different sets of PCMH factors. However, our primary goal in this analysis is to assess whether the specific implementation approach mattered for PCMH performance, rather than to predict the success of a given approach ex ante. Additionally, we are concerned with the comparative success of observed

approaches to implementation, not the full hypothetical space allowed for by the approximately 140-item recognition survey. LASSO-type variable selection approaches might place a substantial weight on PCMH features unlikely to be stressed in the same practices. Moreover, while the models could be designed to consider interactions between factors, these cross-model combinations of factors may never be observed in actual practices. Using the clustering approach sacrifices some ability to identify the influence of individual factors, but ensures that only realistic approaches to PCMH implantation are considered in the analysis.

Moreover, the individual coefficient estimates from such prediction models may not be interpretable in a straightforward way, limiting their usefulness in assessing the impact of components of PCMH recognition. These approaches are engineered to produce accurate predictions rather than stable coefficient estimates, and if substantial collinearities exist between predictors, removing one could lead to large changes in the coefficient estimates of interest while minimally affecting the predicted outcomes. In contexts like ours, where large complementarities in implementation between factors create plausible collinearity, clustering analysis provides an attractive alternative.

One further issue regarding the practice capability data analyzed in this study is that while we were able to observe detailed information about practice attributes post-implementation, we lack any data on the capabilities in place prior to recognition. One assumption we can (confidently) make is that practices neither omitted capabilities which were in place from the PCMH recognition process, nor discarded capabilities which had been in place prior to certification. Acknowledging that in some cases the PCMH recognition process simply catalogs capabilities already in place (rather than catalyzing the introduction of new ones), we can treat our difference-in-differences estimates of the PCMH and cluster effects as lower bound on the true effect, as in the limit, if all capabilities where already in place prior to recognition, its effect should be zero.

2.4.2 Regression Framework

We estimate the impact of practices switching to PCMH status using a generalized difference-in-differences approach. The analytical sample includes only practices which eventually switched to the PCMH model, and practices which had not yet converted at a given point in time serve as controls for the practices switching to PCMH status. The sample is limited to the practices which eventually achieved PCMH recognition and results are robust to inclusion of members in "never adopter" practices, for which detailed recognition data is not available.

However, a regression approach which estimates these effects using a single indicator for the post-treatment period (as in David et al. (2015)) treats the PCMH model as a "black box" and may miss important differences in implementation. Instead, we present the results from the single-indicator approach described above alongside two more flexible alternatives: first, a specification which includes separate "post x treatment" indicators for each of three levels defined by NCQA; and second, which includes "post x treatment" interactions for each of the clusters identified by our hierarchical algorithm. This research design employs an identification strategy which exploits the fact that the transition to PCMH status occurred at different times across primary care clinics, so patient outcomes could be tracked before and after the switch at different points in time and across practices. These effects were estimated using models of the form outlined below:

$$Y_{ijt} = \lambda_t + \mu_i + \beta X_{ijt} + \sum_{g=1}^{G} \delta_g \left(Post\ x\ Cluster_g\right)_{it} + \varepsilon_{ijt} \qquad (2.1)$$

The outcome variable Y_{jit} for patient *j* enrolled at primary care practice *i* during year *t* is either (1) a dichotomous variable tracking whether a patient-year observation includes any hospitalization, or PCP, specialist or ED visit (=1) or no such encounter (=0), or (2) expenditures for professional (physician) services. These expenditures for professional services are important as it captures the total monetary value for encounters with both primary care physicians and specialists. Specialist care is typically less frequent and more expensive than primary care. The PCMH model strives to enable primary care physician to

operate at the top of their license and negate expensive and potentially avoidable downstream specialist care. All regressions with binary outcomes are estimated using linear probability models.[29]

In the tables that follow, the "baseline model" which appears in Panel A includes only the "post" term, indicating that a patient-year is associated with a recognized PCMH practice. In other words, there is only a single PCMH "cluster" of all practices. This "baseline model" ignores the richness of our data and mimics existing studies of PCMH. Panel B extends the analysis to include the "post x level" interactions, and in the models represented in Panel C, we replace them with separate $(Post\ x\ Cluster_c)_{it}$ terms.

The key explanatory variables are the $(Post\ x\ Cluster_c)_{it}$ terms, which are indicator variables capturing each practice's PCMH status during a given year. $(Post\ x\ Cluster_c)_{it}$ equals 1 if an observation was recorded during the first full calendar year following a practice becoming PCMH recognized (or in subsequent years), and if that practice was identified as being a part of cluster c; this indicator is set to zero otherwise. Observations recorded during the transitional year in which the switch to the PCMH model was made were dropped from the analysis in order to account for the challenges of PCMH implementation documented in the literature (Berenson et al., 2008; Harbrecht & Latts, 2012; Kilo & Wasson, 2010). Studies of other health system reforms, including the Massachusetts health insurance reforms, have noted the importance of accounting for such transitional periods in analytical design (Chandra, Gruber, & McKnight, 2011; Joynt, Chan, Orav, & Jha, 2013).[30] λ_t is a year fixed effects term, and μ_i represents patient fixed effects, depending on the specification. All models include these terms in order to account for secular trends and unobserved, time-invariant characteristics of individual patients, respectively. Additionally, the model expressed in equation (2.1) controls for time-varying patient characteristics such as risk score, age, and comorbidities.

[29] These results are qualitatively similar if estimated using a fixed-effects logit model instead, and are available from the authors on request.

[30] Similar results are obtained from models using a "during" variable, which equals 1 in the switch year and zero before and after that year and a variable capturing the fraction of the year in PCMH status.

The error term ε_{jit} represents the remaining, unobserved variation in patient and practice attributes. In all models, standard errors are clustered at the practice level.

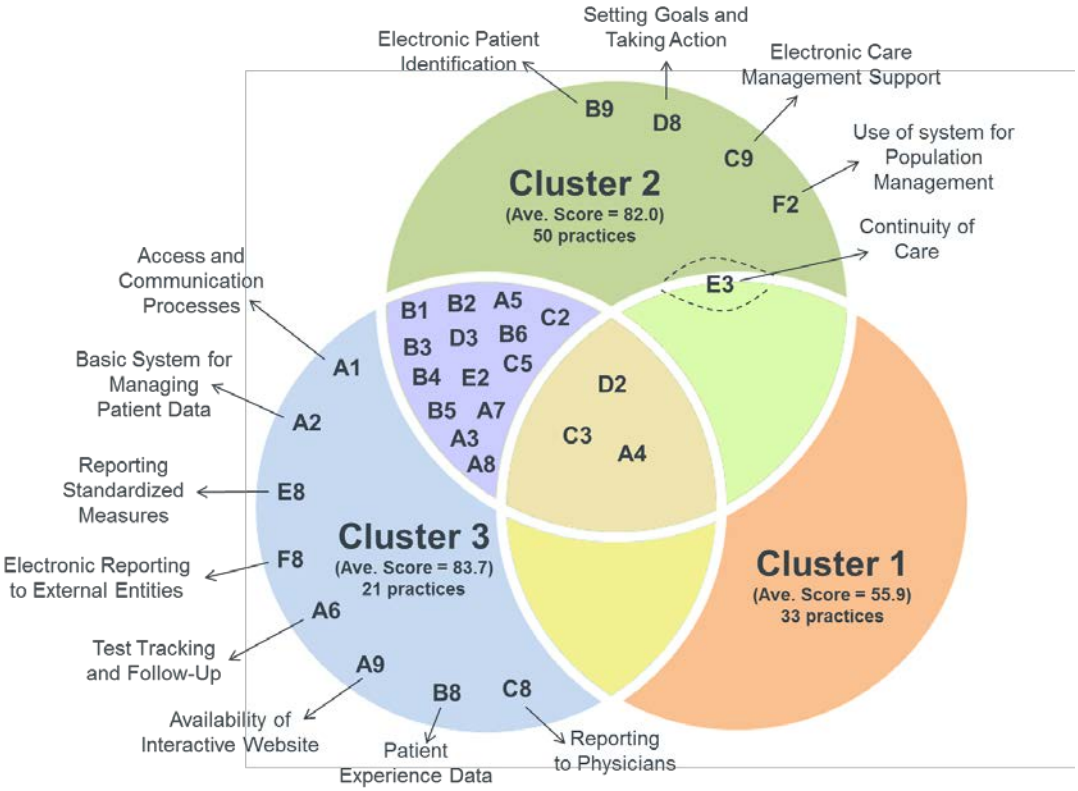
We re-estimate the models on two relevant subsamples and conduct an important robustness check using an alternative pool of controls. First, we limit the sample to the stable panel of continuously enrolled patients, which reduces the risk that selective entry and attrition of patients in PCMH practices is driving the results. We also separately analyze patients with and without documented chronic illnesses, in order to test whether the PCMH model succeeds in its stated objective of better managing the care of patients with complex medical needs. Finally, we re-estimate all of the models using the pool of practices which never achieved PCMH recognition (referred to as "never adopters") as controls.

2.5 Results

2.5.1 Clustering Results

As discussed above, the dendrogram in Figure 2.3 supports the use of a practice typology solution which includes three clusters (labelled 1-3). The three clusters are described using a Venn diagram in Figure 2.4. This diagram was generated by first performing a series of one-way ANOVA tests, which compared the fraction of points within each of the 30 PCMH "elements" that each cluster received. In cases where the clustering explained a significant ($p<0.05$) proportion of the variance in by-element scores, we attempted to identify whether one cluster (or a pair of clusters) dominated the others in terms of implementation. For example, consider element F8 – *Electronic reporting to external entities* - which appears in the Cluster 3 region of Figure 2.4. Practices in Cluster 3 received all possible points associated with element F8 (i.e. 100%), whereas the practices in the Clusters 1 and 2 received averages of 33.3% and 35.5%, respectively. Another example is element B2 – *Electronic system for clinical data* – which received an average score of 98% and 100% in Clusters 2 and 3, while receiving an average score of only 41.7% in Cluster 1. Put differently, nearly all practices in Clusters 2 and 3 have fully implemented B2, while for practices in Cluster 1 there is partial implementation. The score for Cluster 1 can reflect a situation where only two in five

Figure 2.4: Cluster Description Venn Diagram



practices implements this element fully, or where all practices in the cluster have implemented a two-fifths of factors within this element. B2 appears in the area shared by Clusters 2 and 3, which represents the case where the Clusters 2 and 3 dominate Cluster 1. A complete summary of the ANOVA analysis appears in Appendix Table B2.

While most elements are easily assignable to a given location on the diagram, element E3 – *Continuity of care* - was not. Cluster 3 is dominated by the other two clusters, as all its practices had zero points for this element, and at the same time, the scores for Clusters 1 and 2 were not comparable: Cluster 1 received an average of 52% while Cluster 2 received an average of 87%. As a result we placed E3 in-between the area common to Clusters 1 and 2 and the area belonging exclusively to Cluster 1 to highlight the hierarchy across the three cluster as it relates to element E3.

In Figure 2.4, one cluster dominating the other two is represented by an element label appearing in only one circle; if two clusters dominated the third, the element label appears in the bullet-shaped area shared by the two clusters' regions. Finally, the center area includes the elements for which no cluster had a clear advantage over the others (that is, the average element scores were not statistically significantly different across the three clusters). Two general themes of differentiation in PCMH implementation can be attributed to Clusters 2 and 3, based on the labels highlighted in Figure 2.4. Clusters 2 and 3 dominate Cluster 1 in terms of average total points received – 82 and 83.7 versus 55.9. Hence, the relatively few factors recorded for Cluster 1 are also shared by Clusters 2 and 3. The higher score of practices in these two clusters was achieved by (1) recording additional similar factors (presented by specific elements in the area shared by the two clusters) and (2) recording additional factors that differ between Clusters 2 and 3 (presented in areas that belong to a single cluster).

Returning to Table 2.2, we can point to a few descriptive differences between the clusters as well. We observe some variability in age across clusters, with Cluster 2 serving substantially younger patients than 1 or 3. Along with this difference in age across clusters, we see a substantial difference in terms of patient health. The first panel of Table 2.2 points to patients in Cluster 3 practices having the highest risk scores, followed by Clusters 1 and 2. These differences are likely attributable to the age differences across clusters as well as the modest differences in comorbidity prevalence presented in Table 2.2, but as the gradient across clusters persists in the stable panel sample, it is unlikely for these differences to reflect selective entry or attrition of patients.

Cluster 1 is best viewed as a group of practices meeting a set of minimum requirements needed to secure PCMH recognition. As a result, these practices would likely experience smaller improvements in patient outcomes relative to Clusters 2 and 3, when the medical home model is treated as differentiated interventions, rather than a unified approach. Additionally, the two high-performing clusters appear to have different orientations in terms of the subsets of PCMH capabilities they tend to emphasize. In addition to the common elements which distinguish both Clusters 2 and 3 from Cluster 1, Cluster 2 appears to focus

on patient-facing activities. These include population management, electronic support for care management, and improving continuity of care. Cluster 3, on the other hand, appears to have adopted a physician-facing orientation that emphasizes implementation of decision support, data reporting, and adoption of systems for managing and tracking patient data. Most importantly, distinguishing between the practices we have identified as Clusters 2 and 3 is impossible without the detailed certification data used in this study, given their similar PCMH recognition levels and even overall scores.

A second Venn diagram, expressing the differences in emphasis of the 30 elements by NCQA recognition level, appears in Appendix Figure B1. This figure largely reflects the dominance relationships we would expect from the structure of the NCQA rules. We observe a number of elements on which the three levels are not differentiating (including some must-pass elements), a smaller number on which Levels 2 and 3 dominate Level 1, and finally 12 elements for which Level 3 practices distinguish themselves from Levels 1 and 2. There is one outlier – Level 2 practices have an apparent advantage in terms of A4 – *Documenting communication needs*. Additionally, the average score for Level 1 practices exceeds the 50 points needed for Level 2 recognition. The reason for this is the presence of clinics with scores sufficient for achieving level 2 or even level 3 who do not have the required 10 "must pass" elements.

The importance of the Venn diagrams goes beyond illustrating the variation across clusters of NCQA recognition, highlighting two important features. The first is that recognition level may not be very informative. Even practices with similar achievement in total points (information that is only available to the practice) can differ in terms of care orientation and emphasize different productive tasks.[31] Second, outcomes that tend to improve only for patients in one cluster can be subsequently traced back to the underlying elements that separate this cluster from its counterparts. A similar exercise can be performed for elements for which two clusters have an advantage.

[31] Even levels of recognition, which are public information, are not informative. For example, both Cluster 1 and Cluster 2 include practices of all three levels even though the average total score in Cluster 1 is almost 50% higher than that in Cluster 2.

Some of the elements have clear connections to the outcomes of interest. For example, Cluster 2 dominates in terms of element F2, or "Use of a system for population management." Since this covers areas like proactive engagement on preventive and follow-up care, we would expect reductions in use of high-cost, high-intensity services like inpatient care (with the potential for increased use of PCP care to implement this management). Turning to Cluster 3, we find a distinctive emphasis on a number of decision support and reporting areas. Cluster 3 also dominates on element A6, which concerns management of patient test results – these improvements could reduce the need for repeated tests, reducing the number of PCP and specialist visits. Though it offers less sharp predictions about how patient outcomes will change, Cluster 3 also dominates in terms of three performance reporting elements – C8, E8 and F8. It may be the case that improved documentation of treatment and outcomes, as well as the prospect of having these outcomes reported to insurance companies or other external parties, prompt physicians to change practice style.

Generally, we would expect Clusters 2 and 3 to dominate Cluster 1, which has both a lower average overall practice score and no areas of implementation where it has clear advantages over Clusters 2 or 3. We can make some further predictions about relative performance in terms of the specific areas where Clusters 2 and 3 show an advantage. Elements B2 and C2 ("Electronic System for Clinical Data" and "Use of electronic clinical data") do not provide sharp predictions about specific performance areas, but element E2, "Identifying important conditions," would suggest better performance among patients with chronic illness. The elements under Standard 3, "Care Management," would predict increased use of preventive care and medication adherence, potentially with downstream reduction in high-cost, high intensity services like hospitalizations. Element B4, "Self-management support," could potentially reduce overall utilization by substituting self-management for formal contact with the medical system.

2.5.2 Patient Outcome Regression Results

The tables which follow present results from a series of specifications for each outcome of interest. Each is broken down vertically into two sections with three columns. The top row of each table specifies whether

the section refers to the full sample or to the stable panel of patients who were continuously enrolled for the full study period. Within these two groups, each model is estimated first on all patients and then on patients with and without documented chronic illnesses separately (as indicated by the labels in the second row). The coefficient estimates presented in the tables represent percentage point changes in the case of binary outcomes. In the professional expenditures table, the coefficients represent inflation adjusted dollar changes. For ease of interpretation, we have added percent changes (computed by dividing the coefficients by the relevant baseline outcome rate) for all significant estimates to the tables, appearing in italics below the standard errors. Finally, each table is split into three panels: A, which treats the PCMH as a single intervention; B, which estimates the level effects separately; and C, which employs the novel clustering typology. In Panel A results the baselines used are the outcome rates in the full sample and stable patient panel sample. For the results in Panel B, the rates are based on the within-level averages, and in Panel C, the baseline used is the within-cluster rate.

Panel A of Table 2.3 displays the difference-in-differences coefficients for the single PCMH adoption indicator on the probability of having one or more primary care visits in a given year. These results provide some evidence for an overall PCMH effect, with a significant but modest increase driven by changes among the patients without chronic illness. These effects diminish to non-significance when the models are estimated using the stable patient sample. In Panel B, we find similarly weak evidence for any effect of the PCMH levels – the increase is only evident in the Level 3 practices, and then only in the patients without chronic conditions.

Turning to Panel C, we find the results from the specification using separate difference-in-differences terms for each of the three clusters. These coefficient estimates provide substantial evidence that there are heterogeneous impacts of PCMH adoption which depend on the specific constellation of practice improvements put in place. We find consistent evidence of an increase in utilization in Clusters 2 and 3 (3.3-3.8% and 3.1-3.5%, respectively), and a large decrease in Cluster 1 (8.1-11.2%). These results are robust to limiting the analysis to the stable panel, and in Clusters 2 and 3 are stable whether looking at

Table 2.3: The effect of PCMH levels and clusters on utilization of primary care services

	Member Fixed-Effects (All)			Member Fixed-Effects (Panel)		
	All	**Chronic**	**Non-Chronic**	**All**	**Chronic**	**Non-Chronic**
A - Baseline Model						
POST	**0.00714****	0.00605	**0.00992****	0.00148	0.00404	0.00484
	[0.00329]	[0.00512]	**[0.00403]**	[0.00446]	[0.00664]	[0.00555]
	0.88%		***1.27%***			
Observations	370,764	75,320	295,444	130,923	30,055	100,868
R-squared	0.623	0.595	0.630	0.477	0.452	0.488
B - Level x POST interactions						
Level 1 x POST	0.0024	0.0019	0.0025	0.0022	0.0027	0.0022
	[0.00601]	[0.00515]	[0.00503]	[0.00833]	[0.00873]	[0.00969]
Level 2 x POST	0.0018	0.0012	0.0019	0.0032	0.0014	0.0034
	[0.00572]	[0.00894]	[0.00762]	[0.00745]	[0.0105]	[0.00957]
Level 3 x POST	**0.0102*****	0.0027	**0.0164*****	0.0016	0.0027	0.0015
	[0.00337]	[0.00552]	**[0.00437]**	[0.00474]	[0.00696]	[0.00591]
	1.24%		***2.08%***			
Observations	370,764	75,320	295,444	130,923	30,055	100,868
R-squared	0.625	0.608	0.632	0.485	0.480	0.493
C - Cluster x POST interactions						
Cluster 1 x POST	**-0.0620*****	**-0.0832*****	**-0.0499*****	**-0.0756*****	**-0.100*****	**-0.0606*****
	[0.00436]	**[0.00666]**	**[0.00538]**	**[0.00559]**	**[0.00822]**	**[0.00701]**
	-8.05%	***-9.22%***	***-6.79%***	***-9.67%***	***-11.18%***	***-8.15%***
Cluster 2 x POST	**0.0292*****	**0.0360*****	**0.0290*****	**0.0280*****	**0.0322*****	**0.0268*****
	[0.00353]	**[0.00547]**	**[0.00431]**	**[0.00475]**	**[0.00704]**	**[0.00590]**
	3.53%	***3.77%***	***3.64%***	***3.31%***	***3.35%***	***3.29%***
Cluster 3 x POST	**0.0283*****	**0.0334*****	**0.0266*****	**0.0253*****	**0.0293*****	**0.0255*****
	[0.00487]	**[0.00732]**	**[0.00607]**	**[0.00612]**	**[0.00889]**	**[0.00774]**
	3.48%	***3.52%***	***3.45%***	***3.05%***	***3.07%***	***3.24%***
Observations	370,764	75,320	295,444	130,923	30,055	100,868
R-squared	0.623	0.599	0.631	0.479	0.462	0.489

Robust standard errors in brackets; *** p<0.01, ** p<0.05, * p<0.1

All models control for patient demographics (age, age squared, and gender), patient risk score, five comorbid conditions (congestive heart failure, chronic obstructive lung disease, coronary artery disease, asthma and diabetes), year fixed effects and either practice or member fixed effects

patients with or without chronic illness. In Cluster 1, the reductions in PCP utilization are smaller among the patients with chronic illness, but remain large and highly significant.

Table 2.4 presents the results for models estimating the PCMH effect on the probability of having one or more specialist visits in a given year. None of the specifications used yields any significant overall PCMH effects, or significant effects for any of the level-specific treatment indicators. We do find evidence of decreased specialist utilization in Cluster 3 for both the overall and stable panel samples (2.7-3.6%). In both cases, the effect is driven by changes in the patients with chronic illnesses. In the stable panel only, we see a reduction of similar magnitude for Cluster 1, but driven by patients *without* chronic illness. For Cluster 2 only one specification returned a marginally significant result, showing a modest increase in utilization.

Table 2.5 displays the output from regressions predicting the probability of having at least one inpatient hospitalization in a given year. We find significant changes in the probability of an inpatient admission in the two model specifications using the chronically ill subset of patients (both in the stable panel and when all patients are eligible for inclusion) - specifically, a 4.2-5.4% reduction. However, we find no significant changes in the analyses using level-specific treatment indicators. In Panel C, we find evidence of a PCMH effect by cluster, though the results are again mixed. For practices in Cluster 2, we observe a 9.7-11.8% increase in the likelihood of a hospitalization. The point estimate is larger (though only marginally significant, due to sample size reductions) in the stable panel. In Cluster 3 there are marginally significant reductions in hospitalizations overall (6.4-8.7%), and larger and significant reductions among chronically ill patients (11.5-15.8%), with no affect among the non-chronic samples.

Estimates of changes in ED utilization appear in Table 2.6. In panel A, we find a significant effect on ED use in the full sample of chronically ill patients, and a somewhat larger (but less precisely estimated) effect on the patients with chronic illnesses in the stable panel. We see no significant effects in either the level-specific specifications or in in Clusters 2 or 3. However, we observe increased ED utilization in Cluster 1 (ranging from 8.8-13.1%) in five of the six specifications used.

Table 2.4: The effect of PCMH levels and clusters on utilization of specialist care

	Member Fixed-Effects (All)			Member Fixed-Effects (Panel)		
	All	**Chronic**	**Non-Chronic**	**All**	**Chronic**	**Non-Chronic**
A - Baseline Model						
POST	-0.000799	-0.00184	-0.00376	-0.00589	-0.0142	-0.00893
	[0.00393]	[0.00800]	[0.00459]	[0.00546]	[0.0105]	[0.00646]
Observations	370,764	75,320	295,444	130,923	30,055	100,868
R-squared	0.669	0.705	0.654	0.553	0.596	0.533
B - Level x POST interactions						
Level 1 x POST	-0.00105	-0.00767	-0.00953	-0.0123	-0.00249	-0.0113
	[0.00721]	[0.00978]	[0.00575]	[0.0191]	[0.0125]	[0.0783]
Level 2 x POST	-0.0032	-0.00281	-0.00670	-0.00618	-0.00278	-0.00880
	[0.00685]	[0.00425]	[0.00871]	[0.00918]	[0.00274]	[0.0112]
Level 3 x POST	-0.000461	-0.00238	-0.000154	-0.00092	0.0012	-0.00210
	[0.00404]	[0.00876]	[0.00499]	[0.00585]	[0.0113]	[0.00692]
Observations	370,764	75,320	295,444	130,923	30,055	100,868
R-squared	0.669	0.705	0.654	0.553	0.597	0.533
C - Cluster x POST interactions						
Cluster 1 x POST	-0.00564	0.0123	-0.00953	**-0.0151****	0.00331	**-0.0185****
	[0.00521]	[0.0105]	[0.00614]	**[0.00686]**	[0.0131]	**[0.00818]**
				-2.95%		***-4.35%***
Cluster 2 x POST	0.00487	0.0234	0.00218	0.00151	**0.0199***	-0.00116
	[0.00422]	[0.08602]	[0.00492]	[0.00582]	**[0.0112]**	[0.00688]
					2.62%	
Cluster 3 x POST	**-0.0168*****	**-0.0202*****	0.00992	**-0.0185****	**-0.0233****	0.0111
	[0.00582]	**[0.00692]**	[0.0115]	**[0.00751]**	**[0.00903]**	[0.0142]
	-3.44%	**-2.71%**		**-3.58%**	**-3.06%**	
Observations	370,764	75,320	295,444	130,923	30,055	100,868
R-squared	0.669	0.705	0.654	0.553	0.597	0.533

Robust standard errors in brackets; *** p<0.01, ** p<0.05, * p<0.1

All models control for patient demographics (age, age squared, and gender), patient risk score, five comorbid conditions (congestive heart failure, chronic obstructive lung disease, coronary artery disease, asthma and diabetes), year fixed effects and either practice or member fixed effects

Table 2.5: The effect of PCMH levels and clusters on inpatient hospitalization

	Member Fixed-Effects (All)			Member Fixed-Effects (Panel)		
	All	**Chronic**	**Non-Chronic**	**All**	**Chronic**	**Non-Chronic**
A - Baseline Model						
POST	-0.000192	**-0.00864***	-0.000560	0.00166	**-0.00997***	0.00153
	[0.00235]	**[0.00452]**	[0.00229]	[0.00322]	**[0.00527]**	[0.00314]
		-4.24%			***-5.40%***	
Observations	370,764	75,320	295,444	130,923	30,055	100,868
R-squared	0.621	0.681	0.602	0.485	0.562	0.454
B - Level x POST interactions						
Level 1 x POST	-0.00264	-0.00267	-0.00108	0.00357	0.00273	0.00113
	[0.00431]	[0.00943]	[0.00287]	[0.00388]	[0.0118]	[0.00380]
Level 2 x POST	0.00296	0.0188	0.00223	0.001906	0.00279	0.00475
	[0.00409]	[0.0137]	[0.00434]	[0.00542]	[0.0160]	[0.00544]
Level 3 x POST	-0.003316	-0.00952	-0.000617	-0.00123	-0.00127	0.00132
	[0.00241]	[0.00845]	[0.00249]	[0.00345]	[0.0107]	[0.00336]
Observations	370,764	75,320	295,444	130,923	30,055	100,868
R-squared	0.621	0.681	0.602	0.485	0.562	0.454
C - Cluster x POST interactions						
Cluster 1 x POST	0.00323	0.0153	0.000765	0.00631	0.0202	0.00312
	[0.00311]	[0.0101]	[0.00306]	[0.00405]	[0.0124]	[0.00397]
Cluster 2 x POST	-0.000134	**0.0187****	-0.000232	0.00265	**0.0203***	0.00185
	[0.00252]	**[0.00830]**	[0.00245]	[0.00344]	**[0.0106]**	[0.00334]
		9.71%			***11.78%***	
Cluster 3 x POST	**-0.00609***	**-0.0251****	-0.00425	**-0.00799***	**-0.0303****	-0.00219
	[0.00348]	**[0.0111]**	[0.00345]	**[0.00443]**	**[0.0134]**	[0.00439]
	-6.39%	***-11.50%***		***-8.66%***	***-15.80%***	
Observations	370,764	75,320	295,444	130,923	30,055	100,868
R-squared	0.621	0.681	0.602	0.485	0.562	0.454

Robust standard errors in brackets; *** p<0.01, ** p<0.05, * p<0.1

All models control for patient demographics (age, age squared, and gender), patient risk score, five comorbid conditions (congestive heart failure, chronic obstructive lung disease, coronary artery disease, asthma and diabetes), year fixed effects and either practice or member fixed effects

Table 2.6: The effect of PCMH levels and clusters on emergency department utilization

	Member Fixed-Effects (All)			Member Fixed-Effects (Panel)		
	All	**Chronic**	**Non-Chronic**	**All**	**Chronic**	**Non-Chronic**
A - Baseline Model						
POST	0.00501	**-0.00861****	0.00561	0.00653	-0.0131	0.00536
	[0.00325]	**[0.00373]**	[0.00355]	[0.00441]	[0.0111]	[0.00487]
		-4.54%				
Observations	370,764	75,320	295,444	130,923	30,055	100,868
R-squared	0.515	0.571	0.529	0.326	0.411	0.338
B - Level x POST interactions						
Level 1 x POST	0.00119	0.0172	0.00650	0.00104	0.00255	0.00804
	[0.00595]	[0.0106]	[0.00445]	[0.00530]	[0.0133]	[0.00590]
Level 2 x POST	0.00157	0.0200	0.00219	0.00203	0.0222	0.00209
	[0.00565]	[0.0155]	[0.00673]	[0.00742]	[0.0179]	[0.00844]
Level 3 x POST	0.00285	-0.00148	0.00207	0.00236	-0.00431	0.00176
	[0.00334]	[0.00953]	[0.00386]	[0.00472]	[0.0119]	[0.00521]
Observations	370,764	75,320	295,444	130,923	30,055	100,868
R-squared	0.515	0.57	0.529	0.325	0.411	0.338
C - Cluster x POST interactions						
Cluster 1 x POST	**0.0115*****	**0.0178*****	**0.0115****	**0.0119****	**0.0232*****	0.0114
	[0.00430]	**[0.00414]**	**[0.00475]**	**[0.00554]**	**[0.00859]**	[0.00817]
	8.80%	***9.05%***	***9.66%***	***9.69%***	***13.10%***	
Cluster 2 x POST	0.00408	0.00506	0.00376	0.00626	0.0116	0.00409
	[0.00348]	[0.00936]	[0.00381]	[0.00470]	[0.0119]	[0.00519]
Cluster 3 x POST	-0.00176	-0.00767	-0.000449	-0.000879	-0.00304	0.000633
	[0.00480]	[0.0125]	[0.00535]	[0.00606]	[0.0150]	[0.00681]
Observations	370,764	75,320	295,444	130,923	30,055	100,868
R-squared	0.515	0.571	0.529	0.326	0.411	0.338

Robust standard errors in brackets; *** p<0.01, ** p<0.05, * p<0.1

All models control for patient demographics (age, age squared, and gender), patient risk score, five comorbid conditions (congestive heart failure, chronic obstructive lung disease, coronary artery disease, asthma and diabetes), year fixed effects and either practice or member fixed effects

Table 2.7 presents a final set of outcomes – expenditures on professional services. We find no overall effect of the PCMH model, or when using the level-specific interaction terms, on expenditures for professional services. In contrast, in Panel C we observe a consistent pattern in the estimated PCMH effect across both the full and stable panel samples for Cluster 3. In both cases, we see a small but significant reduction in expenditures overall, with a larger effect among the chronically ill patients (and no significant differences among the non-chronically ill). Reductions in expenditures ranged from 6.7 to 12.3% relative to baseline, depending upon sample and specification.

2.5.3 Sensitivity Analysis: "Never Adopter" Practices as Controls and Test for Selective Attrition

The estimates presented in Tables 2.2 through 2.7 use only patients enrolled in practices which eventually achieved PCMH recognition. However, there are a further 1,549 primary care practices operating in the same region which never achieved recognition during the study period. While we lack any of the granular data on specific PCMH capabilities for these practices it is possible to include them in the analysis. While there is no variation in the treatment indicator for these practices, the changes in the dependent variables would provide an alternative source of counterfactual time trends for the PCMH practices. To address this possibility, we re-estimated all of the models with the "never adopters" as controls. The results are similar in magnitude, with identical sign and significance, for all categories of outcomes. These results are included in full as Appendix Table B3.

Additionally, while fixed effects capture time-invariant factors related to practice performance and member characteristics, it may be important to consider bias resulting from evolutionary forces. In particular, improved performance may simply result from attrition of higher utilizers of care. To address this concern, we perform commonly-used variable addition tests (Verbeek & Nijman, 1992; Wooldridge, 2002) and find no evidence of attrition bias.[32]

[32] We use a modified Verbeek and Nijman (1992) variable addition test described in Wooldridge (2002) in which leads and lags of selection indicators are added as regressors. This approach is attractive in this context since it is implementable using a fixed effects specification.

Table 2.7: The effect of PCMH levels and clusters on professional (physician) expenditures

	Member Fixed-Effects (All)			Member Fixed-Effects (Panel)		
	All	**Chronic**	**Non-Chronic**	**All**	**Chronic**	**Non-Chronic**
A - Baseline Model						
POST	-12.21	-1.224	7.381	-42.59	-16.76	-10.39
	[40.26]	[105.1]	[42.33]	[51.37]	[131.1]	[52.25]
Observations	322,539	73,845	248,694	115,601	29,571	86,030
R-squared	0.846	0.820	0.874	0.656	0.705	0.677
B - Level x POST interactions						
Level 1 x POST	-41.48	92.94	-36.23	-56.10	121.4	-27.63
	[50.06]	[128.3]	[53.23]	[61.86]	[156.4]	[63.47]
Level 2 x POST	53.8	40.1	106.5	130.6	314.3	137.1
	[76.37]	[187.6]	[83.44]	[88.02]	[212.0]	[93.48]
Level 3 x POST	-15.21	-121.9	19.73	-57.87	-149.7	-19.02
	[43.57]	[114.9]	[45.71]	[54.77]	[140.6]	[55.64]
Observations	322,539	73,845	248,694	115,601	29,571	86,030
R-squared	0.846	0.820	0.874	0.656	0.706	0.677
C - Cluster x POST interactions						
Cluster 1 x POST	-11.76	175.5	-29.51	-25.95	167.6	-23.37
	[53.64]	[137.5]	[57.05]	[64.74]	[163.3]	[66.58]
Cluster 2 x POST	18.00	18.27	41.90	-22.63	16.63	2.295
	[43.09]	[113.0]	[45.28]	[54.69]	[140.0]	[55.59]
Cluster 3 x POST	**-137.7****	**-360.2****	-83.13	**-138.1****	**-410.1****	-38.08
	[58.82]	**[150.4]**	[62.99]	**[69.82]**	**[175.9]**	[72.19]
	-6.68%	***-9.76%***		***-6.77%***	***-12.32%***	
Observations	322,539	73,845	248,694	115,601	29,571	86,030
R-squared	0.846	0.820	0.874	0.656	0.706	0.677

Robust standard errors in brackets; *** p<0.01, ** p<0.05, * p<0.1

Professional expenditures are inflation adjusted. All models control for patient demographics (age, age squared, and gender), patient risk score, five comorbid conditions (congestive heart failure, chronic obstructive lung disease, coronary artery disease, asthma and diabetes), year fixed effects and either practice or member fixed effects

2.5.4 Parallel Trends Test

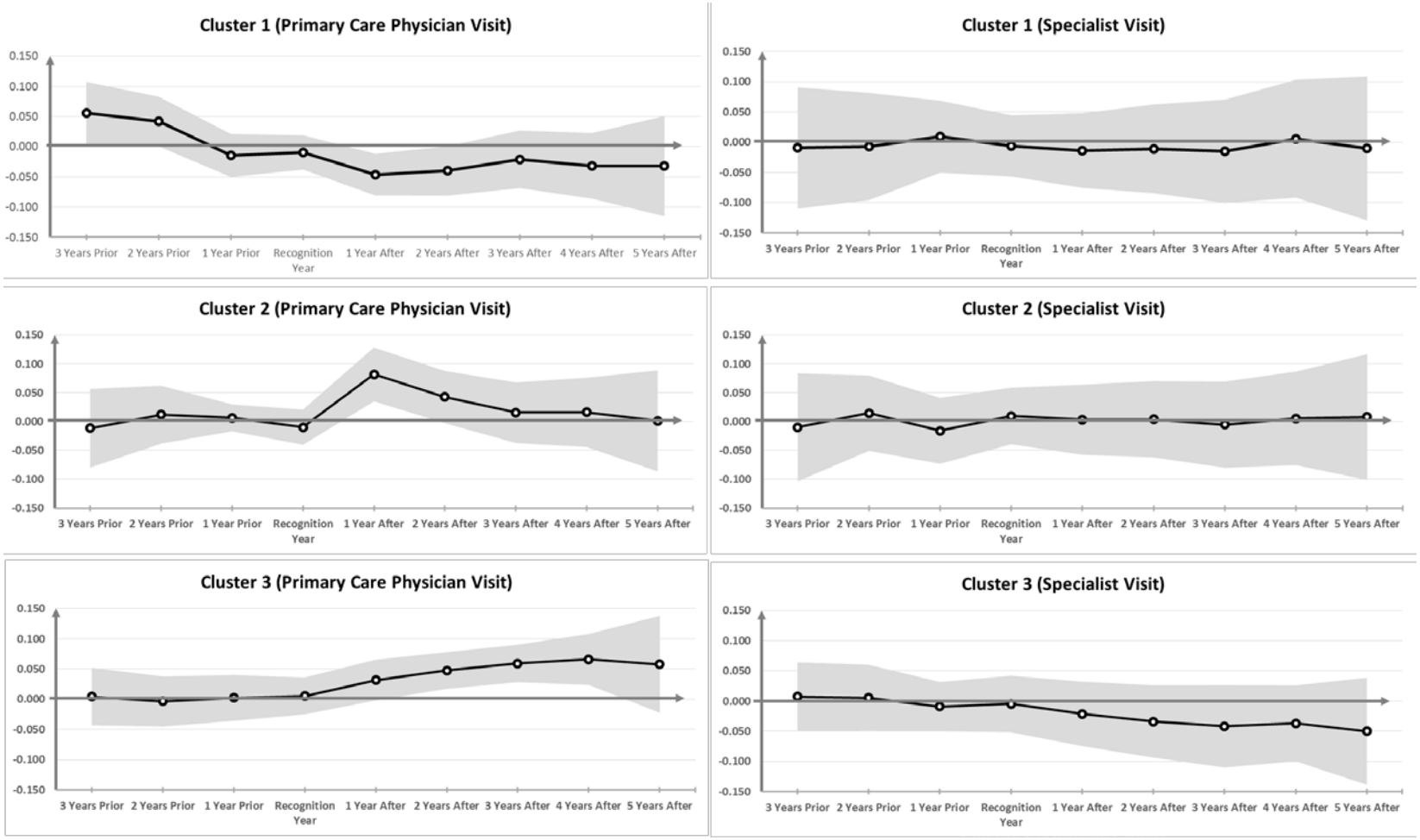
The reliability of results from difference-in-differences analysis depends critically upon confirming the assumption of parallel trends, which requires that the treatment and control practices exhibit similar trends in the absence of treatment. In order to test this assumption, following Autor (2003), we estimate cluster-specific time effects using time-from-recognition year to account for the multiple periods as well as the multiple treatment and control groups in our data. As in the main equation, the outcomes of interest – a dummy for at least one primary care visit in a year and a dummy for at least one specialist visit in a year - are measured at the individual level while the timing coefficients are at the practice level. The specification includes three leads (t-3, t-2, and t-1), the recognition year (t), and five lags (t+1 through t+5). The omitted category is t-4. Hence, δ_Δ is the coefficient on the Δth lead or lag across the three groups, *g*, (clusters or levels).

$$Y_{ijt} = \lambda_t + \mu_i + \beta X_{ijt} + \sum_{g=1}^{G} \sum_{\Delta=-m}^{q} \delta_\Delta \left(D_{jt} \left(\Delta = t - k_j \right) x Group_g \right) + \varepsilon_{ijt} \quad (2.2)$$

While the model has multiple treatment groups (practices) and multiple periods, it is possible to provide a visual inspection for the evolution of practice specific trends as all practices are treated during the sample period, albeit at different years. Figure 2.5 provides this visual by plotting the coefficients of $\delta_\Delta$for the sample including all enrollees.

The shaded area in gray connects the top and bottom of vertical bands representing ±1.96 times the standard error of each point estimate. The figure includes six trends for our two outcomes of interest across the three clusters. A test of the differences-in-differences assumption is $\delta_\Delta = 0$ for all $\Delta < 0$. This assumption holds in the case of specialist visits across all clusters as well for primary care visits for Clusters 2 and 3. In the recognition year, the effects are close to zero, suggesting that changes do not manifest themselves rapidly. With respect to Cluster 1, however, we do find clear pre-trends (that is, the coefficients on the pre-periods are different from zero). This calls into question the reliability of the point estimates for this cluster, and may explain the paradoxical negative impacts of PCMH adoption, such as the reduction in the likelihood

Figure 2.5: Time Effects Pre and Post NCQA Recognition Year for Primary Care Visits and Specialist Visits by Cluster



Note: Time effects are obtained from regressions of mean change on member fixed-effects, year fixed-effects, and dummies for leads, recognition year, and lags (lead t-4 is the omitted category), as described in equation 2.2

of seeing a primary care physician. This is not surprising, as Cluster 1 had the least meaningful adoption of NCQA elements and there is no reason to think that the recognition itself would lead to deterioration in performance.[33] Figure 2.5 indicates that the differences-in-differences strategy seems successful in this context, as the coefficients on the adoption leads are close to zero, showing little evidence of neither an anticipatory response nor implementation preceding recognition.

2.6 Discussion

The reorganization of primary care holds much promise for improving access, quality and cost of care. However, in practice there is little evidence that quality assurance, staffing regulation, and massive investments in technology are bearing fruit. The use of clustering as a first step in analyzing the effects of the PCMH model provides both potentially useful typologies of PCMH practices, as well as a viable strategy for evaluating the different approaches practices have taken in implementation. This approach confers a number of benefits beyond overcoming the obvious dimensionality problem. First, an analysis attempting to estimate the effects of individual factors or elements would likely be complicated by significant collinearity issues: for example, practices would be unlikely to adopt an electronic system for managing patient data, but neither organize nor use it (which are all separate elements). Additionally, the functional form of the primary care "production function" is not precisely understood, and any piecemeal analysis of the PCMH components would ideally consider two-way (and greater) interactions between complementary aspects of the PCMH model, thereby greatly increasing the dimensionality of the problem.

Turning to the estimates of the PMCH effect on utilization, we find that previous analyses which have evaluated the patient centered medical home model as a single, homogenous intervention have missed substantial variation in implementation. Moreover, this heterogeneity may have important implications for patient outcomes. This is true for our naïve PCMH analysis in Panel A of all regression tables, where we

[33] Additionally, the violation of the parallel trends assumption in the case of Cluster 1 practices does not require that we disregard the estimates for Clusters 2 and 3, where the assumption holds.

ignore the richness of our data and treat all recognized practices in the same manner. A summary of the intensity of implementation is the practice score, measured from zero to 100. However, we find that two of our clusters (2 and 3) have roughly equal average scores (82.0 and 83.7) and recognition levels, but differ in their implementation on potentially important elements and exhibit different impacts on utilization and expenditures for their patient population. Moreover, an analysis which uses the greatest level of detail typically available to researchers – practice recognition level – produces largely null results regarding the PCMH effect as well. This suggests that greater detail regarding practice capabilities than is currently available to most researchers will be needed to successfully evaluate the impact of the PCMH and other primary care initiatives.

Additionally, the separate analyses by patients with and without chronic illness point to a need to understand the patient populations most likely to be affected by different patterns of practice improvements under the PCMH model. Our results for specialist utilization, hospitalizations and professional expenditures in particular point to differential impacts in magnitude and significance in the chronically-ill population. These findings are consistent with core principles of population health management, which govern the approach to patient centered care more generally. Our findings suggest that PCMH heterogeneity should not be taken as value-neutral, especially if practices are likely to emphasize PCMH capability areas of relative strength (which provide the easiest path to certification), as our model would suggest. Since not all approaches to implementation are equally effective, the current treatment of the PCMH recognition data as private is problematic, but may not be surprising.

The next question to consider regarding the use of the clustering approach is whether the areas of apparent emphasis have any predictive power with respect to patient outcomes following NCQA recognition, as per the discussion in Section 2.4.1. Cluster 1 can be viewed as a level of basic approval, and may be missing the features that drive improvements in Clusters 2 and 3, explaining their low performance. Cluster 2's patient facing orientation emphasizes features like population health management. This approach could lead to more aggressive outreach including increased screenings, especially for chronically ill patients. In

turn, this would often lead (at least in short run) to a spike in utilization of both specialists and hospitals (but not emergency departments). Finally, Cluster 3 takes a more technology-intensive approach. Emphasizing physician facing activities and efforts to close gaps in treatment, these practices are more likely to see new efficiencies based on ending siloing of patients with complex needs in specialty care, with other providers practicing at the "top of their license."

Practices in Clusters 2 and 3 display increased intensity of primary care services utilization (both in terms of the likelihood of having any visit and the number of contacts), consistent with greater adoption of new capabilities. For cluster 2 we observe more specialist visits and hospitalizations for chronically ill patients, consistent with a story in which outreach leads to greater intensity. On the contrary, in the case of Cluster 3, we find some evidence of an offsetting reduction in hospitalizations. Conversely, we find evidence of reduced PCP utilization in Cluster 1, though this does not translate into significant reductions in spending. Table 2.6 presents some evidence of an increase in emergency department utilization in Cluster 1, pointing to a possible substitution of lower-intensity PCP services for those provided in the ED. Taken together, we present evidence that greater implementation is associated with improved performance – Clusters 2 and 3 dominate on overall PCMH implementation scores, and both provide more primary care. At least in the case of Cluster 3, there is some evidence to suggest that care is provided at lower cost and with fewer hospitalizations.

2.7 Conclusion

Though the patient centered medical home is a leading model for primary care improvement, previous studies in this area (due mostly to data limitations) have not accounted for the substantial heterogeneity in the implementation of the PCMH model, which in turn has significant implications for patient outcomes. Our analysis provides a framework for understanding this heterogeneity, by identifying the subsets of PCMH improvements which different clusters of practices tend to emphasize, and evaluating their differential effects on patients. As others have noted, some degree of flexibility in implementation may

indeed have been essential to encourage uptake of a new model of care like the PCMH. However, as these policies mature (and as funds are directed to promoting implementation), identifying the elements of the PCMH which are driving improvements in patient outcomes becomes an essential task. Otherwise, practices may simply implement the lowest-cost capabilities needed to reach recognition, rather than the highest-value elements from a societal standpoint.

As alternative payment models like Accountable Care Organizations (ACOs) grow in prominence, it will be important not only to offer primary care practices incentives to control costs, but also to provide them with the evidence-based guidelines for population health management that make these cost-savings possible. In this sense, the PCMH model may best be viewed as a complement to other care coordination and reorganization approaches, rather than as a substitutes, and it will be essential to offer primary care providers clear insights about which improvements can reduce costs and utilization of expensive, high-intensity healthcare services.

Chapter 3

The Impact of Patient-Centered Medical Homes on Medication Adherence

with Guy David, Aaron Smith-McLallen, Spencer Luster and Ravi Chawla

3.1 Introduction

Accreditation, certification and recognition of healthcare facilities, while often voluntary, are designed to assess an organization's level of performance in relation to established quality, service and safety standards. For patients, these processes may help resolve information asymmetries pertaining to care quality. The demand for such information is high, as indicated by the proliferation of quality reports and other transparency initiatives since the introduction of cardiac surgery report cards in the early 1990s (Dranove, Kessler, McClellan, & Satterthwaite, 2003). Even though meeting accreditation standards claims resources from healthcare organizations, the process is embraced by a variety of organizations in the US (Christianson, Volmar, Alexander, & Scanlon, 2010).

Recognition is carried out by third party organizations like the National Committee for Quality Assurance (NCQA), The Joint Commission (TJC), and others. These organizations are tasked with defining, measuring and scoring performance by evaluating a large number of attributes. For example, the Centers for Medicare and Medicaid Services Hospital Compare rankings evaluate facilities on 57 dimensions of performance (Medicare.gov, 2016). Making these evaluations usable – particularly in the case of consumer-facing reports – often entails collapsing numerous, complex dimensions of performance into one or a few summaries. In many cases, a simple pass/fail designation is the final product.

The resulting accreditation may still serve as a signal of high performance that reveals quality and reduces the information asymmetry faced by consumers. However, the various attributes considered in the accreditation process are typically tied to different performance dimensions. As a result, the use of a binary indicator of accreditation will obscure meaningful variation in implementation choices and make it difficult to assess its impact on specific performance dimensions. This problem is particularly significant for two reasons. First, different players in the healthcare ecosystem emphasize different dimensions of performance. This has the tendency to bring more diffuse, uncorrelated elements into an accreditation rubric, adding noise to the accreditation signal. Additionally, resource constrained organizations interested in obtaining accreditation can take advantage of the flexibility in implementation that is built in to most accreditation processes, exploiting complementarities in one dimension of performance or another. This can lead to relative neglect of some dimensions, and heterogeneity in implementation is likely to result in heterogeneity in performance.

This study focuses on one specific recognition effort: the patient centered medical home (PCMH) model. The PCMH is a widely-promoted approach to restructuring primary care, with a particular focus on better addressing chronic health conditions. The PCMH model aims to provide care for the "whole person," with an emphasis on population health management and improved coordination of care. Particularly, this entails the adoption of healthcare information technology to manage medical records and increase access to evidence-based care guidelines, as well as expanding options for communication with providers. A more extensive treatment of the history of the PCMH model has been provided by the health services literature (Friedberg et al., 2009; Kilo & Wasson, 2010).

To achieve recognition as a medical home, practices start by reviewing the requirements for a given recognition standard and assessing of capabilities currently in place. Practices then create a plan to adopt any missing features, and finally submit documentation for recognition. Since 2008, the National Committee on Quality Assurance (NCQA) has recognized over 14,000 primary care practices as achieving

PCMH status. The complete recognition checklist specifies over 100 areas of performance on which practices are evaluated.

One important clinical issue to which practices can choose to devote greater or lesser attention in the PCMH recognition process is medication adherence, or the extent to which patients follow prescribed medication orders. Adherence has a significant impact on health outcomes with more adherent patients experiencing better outcomes (Baroletti & Dell'Orfano, 2010; Koster, Philbert, Winters, & Bouvy, 2015; Roumie et al., 2011) and lower medical costs (Sokol et al., 2005) than those with suboptimal adherence. Estimates of the cost of non-adherence to the US health system range from over 100 billion annually in hospital spending (Osterberg & Blaschke, 2005) to more than 290 billion annually overall (The New England Healthcare Institute, 2009). Furthermore, there is evidence that increased drug costs associated with increased adherence can be offset by savings elsewhere, especially in hospital inpatient and emergency department expenditures (Roebuck, Liberman, Gemmill-Toyama, & Brennan, 2011).

Medication adherence is particularly important in chronic disease management (Baroletti & Dell'Orfano, 2010; Franks, Burton, & Simpson, 2005; Khdour, Hawwa, Kidney, Smyth, & McElnay, 2012; Krass, Schieback, & Dhippayom, 2015; Monane, Bohn, Gurwitz, Glynn, & Avorn, 1994; Osterberg & Blaschke, 2005). There are multiple reasons patients may be non-adherent, including the costs of medications (Gadkari & McHorney, 2012; Mojtabai & Olfson, 2003), forgetfulness (Franks et al., 2005; Gadkari & McHorney, 2012), inadequate health literacy (Lindquist et al., 2012), and low levels of knowledge about the medications prescribed and the disease being treated (Khdour et al., 2012; Koster et al., 2015; Tae et al., 2016). A variety of interventions to improve medication adherence have been proposed (Nieuwlaat et al., 2008), including a number of patient-centered approaches (Mishra, Gioia, Childress, Barnet, & Webster, 2011; Roumie et al., 2011).

This study is the first to pair detailed element-level data from the NCQA to patient claims data to study the impact of adoption of adherence related measures on improved adherence. Even without the ability to study

specific recognition elements, research on medication adherence in the context of medical homes has shown positive effects. For example, higher adherence among Medicaid medical home enrollees was found for children with asthma (M.E. Domino, Humble, Lawrence Jr., & Wegner, 2009), adults with schizophrenia and major depressive disorder (M.E. Domino, Wells, & Morrissey, 2015), and adults with hypertension, diabetes and hyperlipidemia (Beadles et al., 2015). Outside of Medicaid, adherence increased in two pilot tests of a medical home model in the Geisinger Health System (McCarthy, Mueller, & Wrenn, 2009) and recently, in a national sample of patients and providers (Lauffenburger et al., 2017). Finally, a recent study of patients with severe mental illness found that the estimated PCMH effect on a range of clinical outcomes (including adherence) was robust to a variety of statistical techniques (Domino, Kilany, Wells, & Morrissey, 2017).

Given the heterogeneity in how practices achieve PCMH recognition it is possible that the impact of PCMH recognition on adherence is sensitive to the adoption of recognition elements that deal explicitly with medication adherence. In this paper we use a novel and detailed dataset describing the presence of specific PCMH capabilities related to adherence at the practice level to assess whether individual practice capabilities or specific combinations of adherence-related PCMH components had a particular effect on adherence overall and within certain therapeutic classes.

3.2 Data

3.2.1 Practice Data

The following analyses study PCMH adoption as defined and recognized by the NCQA; as such, any reference to the PCMH model refers to the specific guidelines laid out by NCQA. In addition to the NCQA recognition level (I, II, or III) and the date of recognition, we obtained detailed practice-level data consisting of PCMH standard, element, and factor[34] level scoring considered by NCQA for recognition. Data release

[34] That is, the most detailed level of information collected and used in the 2008 PCMH recognition process.

authorizations were sent to 280 practices that had received PCMH recognition between 2008 and 2012 and were contracted through HMO arrangements with Independence Blue Cross. We obtained authorization from 134 practices, including 27 pediatric practices and 107 practices treating primarily adult populations. To take advantage of the detailed recognition data and avoid potential selection issues resulting from including "never adopters" as controls, the analytical sample includes only practices that switched to the PCMH model during our sample period.

3.2.2 Patient Data

Our dataset includes all commercial HMO and Medicare Advantage (MA) patients enrolled in these 134 PCMH practices with continuous enrollment between January 2008 and December 2013. We use these patients because all HMO and MA enrollees are required to identify a primary care provider, whereas PPO enrollees are not. This identification of a provider of record allows for allocation of patients to specific practices. Additionally, enrollees must have had a pharmacy benefit in order to obtain the data used to calculate the adherence outcomes measures.

We omitted patient data from the year that their providers' practice was first recognized as a PCMH (Berenson et al., 2008; David et al., 2015; David, Saynisch, & Smith-McLallen, 2018; Harbrecht & Latts, 2012; Kilo & Wasson, 2010). Since adherence is measured at the patient-year level, removing the recognition year allows us to avoid attributing adherence to the fraction of the year before or after PCMH recognition. Models including the recognition year data and employing separate controls for that year yield similar results, but are not as conservative as our approach here. The data used for this study span 6 years, representing a significant improvement over some earlier evaluations of the PCMH model. Our final sample consists of 66,024 unique patients and 147,496 total patient-years.

3.2.3 Outcomes Measures

The outcome of interest in this study is medication adherence, measured using the Proportion of Days Covered (PDC) metric, which is defined as the ratio of days in which the patient possessed all prescribed

drugs to total calendar days during which a patient was eligible for inclusion in the study. PDC values of 80% or higher are frequently categorized as "adherent" in the literature. However, this approach sacrifices information by only presenting whether a patient was above or below some cut-off. Additionally, Roebuck, Kaestner, & Dougherty (2018) point to recent work indicating that even if an "optimal" PDC threshold existed, it would likely vary by pharmaceutical agent and treatment context. Following these insights, we report the continuous PDC value, summarized at the annual level for each patient, rather than a binary indicator for adherence. The dataset used to conduct the following analyses employ the American Hospital Formulary Service (AHSF) classification system to identify the drugs used by any given patient. The full list of codes and the mapping of individual drugs to the categories used in this paper appear in Appendix Table C1. Values for PDC were calculated for all members recorded as using the drugs listed in Appendix C1, regardless of diagnosis codes.

Table 3.1 provides some descriptive statistics concerning the population of interest, including the numbers of patient-year observations with any adherence data and of patients on medications within several common therapeutic classes. The top row of the table provides data on the sample size for each set of analyses. Overall, we were able to observe adherence rates for 147,496 patient-year observations, though this number varies by category (from a low of 6,148 for insulin to 68,290 for lipid control medications). The average age for patients with prescription medication records is 56.7 years, with asthma/COPD patients substantially younger, given the large fraction of pediatric asthma patients. Table 3.1 also provides rates of chronic illnesses of interest in the sample. Finally, the last two rows of the table display the rates of adherence (measured by PDC) in the study sample. The overall rate of adherence is nearly 74%.[35] Adherence ranges from a low of 53% for asthma and COPD medication to a high of 81% in the case of beta blockers.

[35] It is important to keep in mind that the rate shown here is average total adherence, rather than the proportion of patients who are greater than 80% adherent as is sometimes reported.

Table 3.1: Descriptive Statistics – Patients, Overall and By Drug Category

	Overall	Oral Diabetes	Lipid	Insulin	Beta Blockers	Asthma COPD	Anti-depressants	ACE Inhib.
Num of Obs.	147,496	20,639	68,290	6,148	60,016	20,868	48,534	58,079
Age	56.67	64.09	64.78	58.95	65.85	43.32	51.45	64.58
	(19.52)	(14.41)	(13.46)	(18.22)	(15.75)	(24.54)	(18.57)	(14.09)
Female	0.58	0.53	0.52	0.51	0.57	0.60	0.70	0.54
	(0.49)	(0.50)	(0.50)	(0.50)	(0.50)	(0.49)	(0.46)	(0.50)
Risk Score	4.46 (9.65)	5.97 (11.46)	5.45 (10.60)	10.17 (18.18)	6.84 (12.61)	4.40 (10.15)	4.63 (10.30)	5.48 (10.73)
CHF	0.04 (0.19)	0.07 (0.25)	0.06 (0.23)	0.13 (0.34)	0.08 (0.27)	0.04 (0.20)	0.03 (0.18)	0.06 (0.25)
COPD	0.06 (0.23)	0.07 (0.25)	0.07 (0.26)	0.09 (0.28)	0.08 (0.27)	0.17 (0.38)	0.05 (0.23)	0.07 (0.25)
CAD	0.12 (0.33)	0.20 (0.40)	0.22 (0.41)	0.26 (0.44)	0.25 (0.43)	0.08 (0.27)	0.08 (0.27)	0.19 (0.39)
Asthma	0.14 (0.35)	0.12 (0.33)	0.09 (0.29)	0.16 (0.36)	0.09 (0.29)	0.76 (0.43)	0.11 (0.32)	0.10 (0.30)
Diabetes	0.19 (0.39)	0.94 (0.23)	0.29 (0.45)	1.00 (0.07)	0.25 (0.43)	0.11 (0.32)	0.13 (0.34)	0.30 (0.46)
Adherence (PDC)	73.51 (25.27)	79.53 (24.01)	78.59 (23.86)	67.07 (25.04)	81.35 (23.58)	52.85 (30.30)	70.19 (28.21)	80.59 (22.94)

Note: standard errors appear in parentheses

Table 3.2 presents descriptive statistics on the types of practices in our sample (family practice, internal medicine, pediatrics or nurse practitioner), the total PCMH recognition points acquired by the averagepractice, and the presence of the following six adherence-related PCMH capabilities:

1. Standard 2, Element D: **Organizing Clinical Data** - The practice uses the following electronic or paper-based charting tools to organize and document clinical information in the medical record

2. Standard 2, Element F, Factor 3: **Use of System for Population Management** - The practice uses electronic information to generate lists of patients on a particular medication and take action to remind patients or clinicians proactively of services needed
3. Standard 3, Element D: **Care Management for Important Conditions** - For the three (practice-identified) clinically important conditions, the physician and nonphysician staff use specified care management support guidelines
4. Standard 5, Element A: **Electronic Prescription Writing** - The practice uses an electronic system to write prescriptions using either a stand-alone system with either print capability at the office or the ability to send a fax or electronic message to a pharmacy, or one that is linked to patient-specific demographic and clinical information.
5. Standard 5, Element B: **Prescribing Decision Support—Safety** - Clinicians in the practice write prescriptions using electronic prescription reference information at the point of care, including alerts and information on interactions, allergies, dosing, and duplication
6. Standard 5, Element C: **Prescribing Decision Support—Efficiency** - Clinicians engage in cost-efficient prescribing through use of an electronic prescription writer with general automatic alerts for different choices including generics, and/or one connected to payer-specific formulary that automatically alerts the clinician to alternative drugs, including generics.

These capabilities are represented by elements/factors which in consultation with NCQA were identified to be most closely relate to adherence. Each element was scored for purposes of PCMH recognition either using a percent-achievement standard, or a set of four qualitative categories. Note that there is variation in the propensity to implement different recognition components. For example, nearly all of the practices had element 2D in place, though roughly half (51 %) had implemented Factor 2F3.

In addition to the 134 practices included in the analyses which follow(the “responders”), Table 3.3 presents summary statistics for two relevant comparison groups: first, for the 146 practices which did not respond

Table 3.2: Descriptive Statistics - Practices

Attribute	# Practices	Proportion
Family Practice	60	44.78%
Internal Medicine	45	33.58%
Pediatrics	27	20.15%
Nurse Practitioner	2	1.49%
2D: Organizing Clinical Data	128	95.52%
2F3: Use of System for Population Management	68	50.75%
3D: Care Management for Important Conditions	124	92.54%
5A: Electronic Prescription Writing	88	65.67%
5B: Prescribing Decision Support - Safety	115	85.82%
5C: Prescribing Decision Support - Efficiency	110	82.09%

Attribute	# Practices	Mean	S.D.
Total Points	134	74.50	15.02

to the request for data sharing ("non-responders") and 1,805 practices which never achieved PCMH recognition during the study period ("never-adopters").

In terms of practice size and patient characteristics (age, sex, overall risk score and rates of key chronic illnesses), the responders and non-responders are statistically indistinguishable. They do differ, however, on baseline rates of adherence with respect to three of the seven categories of medications examined in this study: lipid controlling agents, oral diabetes treatments and anti-depressants. While non-responders had lower rates of adherence for three medication categories, there was no statistically significant difference in overall adherence rates.

In contrast to the similarity between the responder and non-responder practices, the practices which never achieved PCMH recognition look very different. For example, members in never-adopter practices are older, sicker, and exhibit lower adherence rates. Additionally, never adopter practices have far fewer patients covered by the commercial insurer who provided our data compared with the eventual adopters (80 versus 234 members per practice). This could be because patients enrolled in the insurer whose data we

Table 3.3: Summary Statistics for Responders, Non-Responders and Never-Adopters

	Responders (N=134)	Non-responders (N=146)	Never Adopters (N=1,805)
Average Members per Practice	234.07 (233.55)	215.87 (249.98)	80.31 *** (115.46)
Adherence			
Overall (PDC)	72.78 (11.10)	72.45 (11.40)	70.73 (14.76)
ACE	78.82 (11.22)	77.35 (11.05)	75.09 *** (13.46)
Beta Blockers	76.38 (13.16)	76.01 (13.46)	75.72 (15.35)
Insulin	69.19 (13.79)	66.52 (12.66)	63.90 *** (16.82)
Lipid	77.28 (9.45)	74.10 ** (12.70)	72.97 *** (13.83)
Oral Diabetes	78.17 (9.97)	75.10 ** (12.46)	74.57 ** (15.78)
Asthma / COPD	51.51 (10.54)	52.06 (11.91)	49.95 (17.28)
Antidepressants	69.43 (8.72)	66.14 *** (11.71)	64.68 *** (16.80)
Demographics			
Average age	50.26 (19.30)	51.31 (19.43)	54.25 ** (18.20)
Risk	3.73 (1.82)	3.68 (1.76)	4.21 (4.55)
Female	0.55 (0.13)	0.56 (0.14)	0.54 (0.21)
Chronic Cond.			
Asthma	0.22 (0.21)	0.21 (0.21)	0.17 ** (0.21)
CAD	0.09 (0.07)	0.09 (0.07)	0.11* (0.13)
CHF	0.01 (0.09)	0.01 (0.08)	0.04 * (0.20)
COPD	0.05 (0.06)	0.05 (0.06)	0.04 ** (0.04)
Diabetes	0.18 (0.38)	0.14 (0.35)	0.20 *** (0.40)

Note: Values for Age, Risk Score and Adherence are means; values for Female and chronic illness categories are proportions. Standard errors appear below in parentheses.
* p<0.1, ** p<0.05, *** p<0.01, indicating significant difference from responders

accessed make up a substantially smaller share of the practice, and/or because these practices have smaller overall patient panels. In both cases, PMPM reimbursement incentives to become a PCMH are diluted.

3.3 Methods

Our analysis of the PCMH model's impact on adherence begins by first ignoring the richness of our data to mimic the standard approach to evaluating the medical home, using only an indicator of whether a practice had achieved PCMH recognition as the treatment of interest. Next, using the detailed, practice-level data on adherence-related capabilities, we collapsed each feature into a binary variable separating weak or no implementation from strong implementation (see Appendix Table C2 for a detailed description of the scoring used in these guidelines and our coding of the scoring).

We proceed with two distinct approaches. In the first analysis, we remain agnostic to whether any of the six PCMH components is more important than another, and assess how many adherence-related components a given practice has in place. This strategy resulted in practices being partitioned into seven mutually exclusive groups. While this approach ignores the actual components implemented, it features a natural gradient from zero to all adherence-related components.

In the analyses that follow, the sample is limited to practices which achieved PCMH recognition at some point during the study period. In other words, the control group to which adopters are compared are practices which were not yet recognized as PCMH at any given point in time, but would do so by the end of 2013. Practices never achieving recognition were excluded from the sample entirely. We apply these criteria for two reasons: first, the detailed recognition data used to test the impact of medication-adherence relevant capabilities are only available for practices which underwent the recognition process; and second, it avoids comparing "adopters" and "never adopters," who may exhibit fundamental unobserved differences in resources or organization that could bias our results.

We estimate the impact of practices switching to PCMH status using a generalized difference-in-differences approach, consistent with the formulation of the generalization in Bertrand, Duflo, & Mullainathan (2004),

Hansen (2007), or Imbens et al. (2009) This research design employs an identification strategy which exploits the fact that the transition to PCMH status occurred at different times across primary care clinics, so patient outcomes could be tracked before and after the switch at different points in time and across practices. The analytical sample includes only practices which eventually switched to the PCMH model, and practices that had not yet converted at a given point in time serve as controls for the practices switching to PCMH status. The sample is limited to the practices which eventually achieved PCMH recognition for two reasons: first, the detailed recognition data is only available for these practices; and second, to address potential selection issues which would result from including "never adopters" in the control group. This research design employs an identification strategy which exploits the fact that the transition to PCMH status occurred at different times across primary care clinics, so patient outcomes could be tracked before and after the switch at different points in time and across practices. Each of the models presented is estimated using ordinary least squares, and takes the following general form:

$$Y_{ijt} = \lambda_t + \mu_i + \beta X_{ijt} + \sum_{c=1}^{C} \delta_c \, (Post \; x \; Treatment_c)_{it} + \varepsilon_{ijt} \qquad (3.1)$$

where δ_c is the parameter of the interest – the effect of the PCMH "treatment." In the baseline specification, Post x Treatment is a single variable which indicates whether a patient-year observation is associated with a practice that has achieved PCMH recognition or not. In the subsequent models, we use groups of indicator variables providing more detailed descriptions of PCMH implementation instead of the binary indicator for adoption. The second specification uses ordinal indicators for whether a practice has implemented 0 or 1, 2 or 3, and 4, 5 or all 6 adherence-related capabilities. The third specification uses all 20 combinations of the six adherence-related components which appear in our sample, and we report coefficient estimates for the 16 permutations which were present in more than one practice and therefore can be identified using our empirical strategy. These models assess whether one particular configuration has particularly beneficial impacts on adherence.

In the main analyses, the outcome variable Y_{jit} for patient *j* enrolled at primary care practice *i* during year *t* is the PDC adherence measure. Observations recorded during the transitional year in which the switch to the PCMH model was made were dropped from the analysis in order to account for the challenges of PCMH implementation documented in the literature (Berenson et al., 2008; David, Gunnarsson, Saynisch, Chawla, & Nigam, 2015; David, Saynisch, & Smith-McLallen, 2016; Harbrecht & Latts, 2012; Kilo & Wasson, 2010). λ_t is a year fixed effects term, and μ_i represents patient fixed effects. All models include these terms in order to account for secular trends and unobserved, time-invariant, characteristics of individual patients. Additionally, the model expressed above controls for time-varying patient characteristics such as risk score, age, and comorbidities. The error term ε_{jit} represents the remaining, unobserved variation in patient and practice attributes. Standard errors are clustered at the practice level. In the tables that follow, only the coefficients of interest (that is, the percentage point estimates from the *Post x Treatment* difference-in-differences indicator) are reported, though every model includes the full slate of control variables described above.

3.4 Results

3.4.1 Medication Adherence

The results from the regression analyses described in the previous section are presented in Tables 3.4 and 3.5. In both tables, each column represents a separate model with the outcome of interest specified in the top row. Table 3.4 includes two panels, each representing a set of regressions using differently constructed PCMH treatment variables. In the first panel of Table 3.4, the coefficient estimates represent the increase in adherence caused by a practice being recognized as a PCMH. To reiterate, these "baseline" models use the standard approach from the PCMH literature, treating the PCMH model as a single treatment irrespective of the specifics of implementation. In the second panel, our models replace the single Post x PCMH variable with a series of "Post x Treatment" indicators capturing the number of adherence-relevant capabilities implemented.

Table 3.4: Impact of PCMH on Medication Adherence (PDC) – Baseline and Number of Elements

	Adherence (PDC) by Medication Category							
	Overall	**Diabetes**	**Lipid**	**Insulin**	**Beta Blockers**	**Asthma/ COPD**	**Antidep.**	**ACE**
Model 1: Baseline								
Post x PCMH	2.114*** [0.285]	2.888*** [0.770]	1.601*** [0.415]	0.262 [1.455]	2.235*** [0.422]	2.533** [0.993]	3.568*** [0.625]	2.347*** [0.460]
Model 2: Elements Post X								
0 or 1 Elements	-4.691 [19.64]	25.17 [21.01]	-0.683 [19.77]	-5.911 [4.32]	-9.347 [18.35]	5.377 [6.112]	-5.031 [8.791]	29.92 [18.14]
2 or 3 Elements	1.478** [0.575]	2.322 [1.623]	0.988 [0.843]	2.612 [2.865]	2.041** [0.807]	2.505 [2.039]	1.303 [1.342]	0.698 [0.930]
4 Elements	2.211*** [0.445]	2.901** [1.208]	1.875*** [0.652]	2.163 [2.460]	2.146*** [0.638]	3.391** [1.638]	2.572** [1.006]	2.733*** [0.700]
5 Elements	2.336*** [0.315]	2.861*** [0.851]	1.611*** [0.456]	-1.297 [1.625]	2.488*** [0.464]	2.614** [1.149]	4.099*** [0.702]	2.645*** [0.501]
All 6 Elements	1.837*** [0.357]	3.152*** [0.980]	1.645*** [0.523]	1.464 [1.851]	1.857*** [0.530]	2.050* [1.229]	3.777*** [0.785]	1.998*** [0.587]
Obs.	147,496	20,639	68,290	6,148	60,016	20,868	48,534	58,079
R^2	0.778	0.746	0.739	0.781	0.092	0.842	0.791	0.73

Note: Coefficients presented are percentage point changes in PDC. Standard errors are clustered at the practice level and appear in brackets. Each model includes the full set of controls listed in the text, but only the DID coefficients are reported. * p<0.1, ** p<0.05, *** p<0.01

Our baseline model indicates that PCMH implementation led to significant increases in adherence, both overall and in nearly all of the prescription drug categories of interest (except for insulin, where sample size issues may result in inadequate statistical power). PCMH recognition led to a greater than 2 percentage point increase in overall adherence (a 2.9% increase relative to the average level of overall adherence in the sample as a whole); effect sizes for the individual categories ranged from 1.6 percentage points for statins to nearly 3.6 percentage points for antidepressants.

3.4.2 Adoption of Adherence-Related PCMH Capabilities

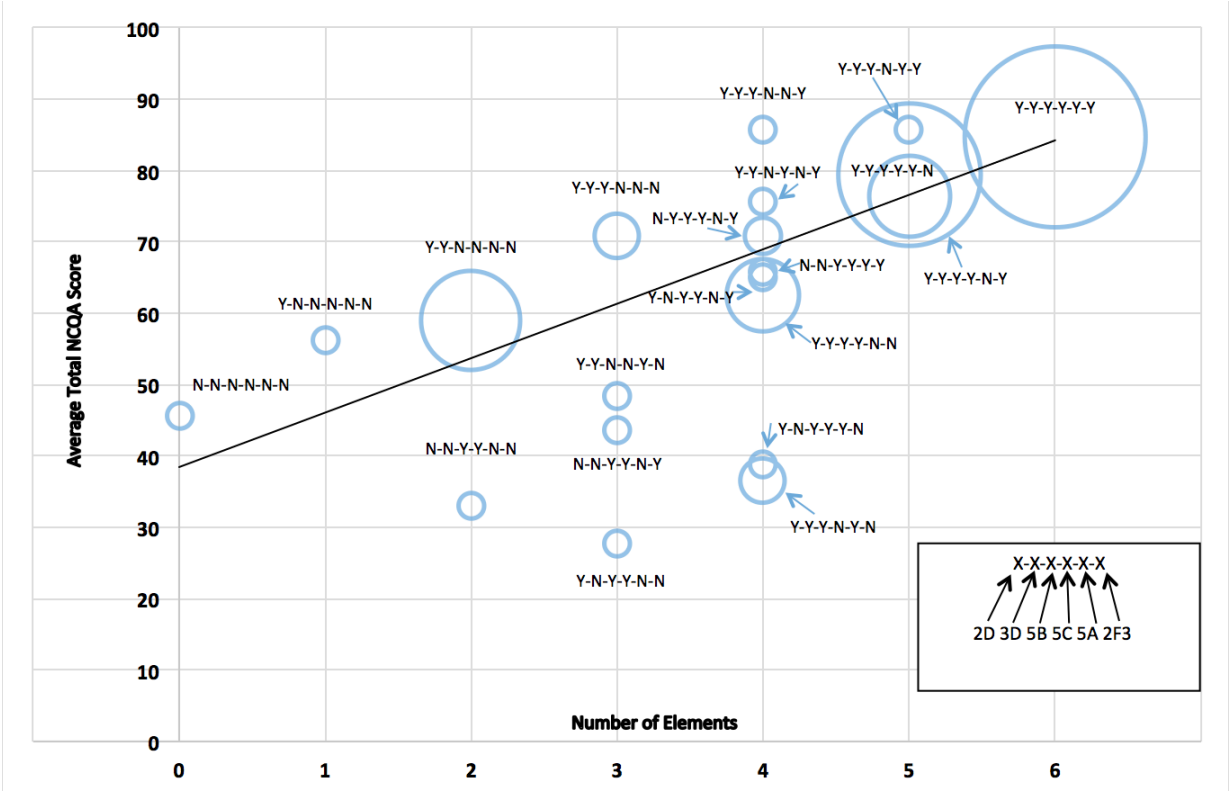
The bottom panel of Table 3.4 suggests no effect of implementation of 0 or only 1 of the identified PCMH elements, and significant increases in overall adherence and adherence to beta blockers in the "2 or 3 elements" category. However, for practices with 4-6 of the identified PCMH elements, we find significant increases across the board (with the exception of insulin adherence) – mirroring the results from the "baseline" model.[36]

Figure 3.1 summarizes the range of patterns of adoption of the adherence-related PCMH capabilities mentioned in the previous section. The numbers along the x-axis indicate how many of the six items of interest are in place for each practice. As shown in the inset, each possible configuration of inputs for a given total is summarized with a string of six yes/no indicators. The radius of each circle indicates the relative number of practices with each configuration. Finally, the y-axis captures the average total PCMH points scored by practices in each cell. We can observe that while number of adherence-related components and total score are positively correlated, the relationship is far from deterministic. In other words, even given more granular summaries of PCMH implementation (such as level or total points), analyses not using the detailed implementation data may miss important variation in how primary care practices are operating following PCMH recognition.

There are 64 possible combinations of the six adherence-related PCMH elements, but only 20 combinations appear in our sample. Moreover, four permutations appear in only a single practice; because the models presented include time fixed effects, the PCMH effect is subsumed by the time effect. As a consequence, we present estimates for the 16 combinations for which we are able to estimate configuration-specific effects.

[36] One potential concern here is that greater adoption of adherence-related elements proxies for more PCMH adoption in general, and it is the comprehensiveness of adoption that is driving adherence and not (or not only) the adherence-related measures per-se. To address this, we repeated the models presented in Table 3.3 with an additional control for total score net of adherence-related elements. The results are largely unchanged, suggesting that improvements in adherence rely predominantly on adherence-related measures and is independent of adoption of other measures.

Figure 3.1: Different Configurations of the Six Medication Adherence Elements and Total NCQA Score



These results, presented in Table 3.5, parallel those presented in the bottom panel of Table 3.4 – this is to be expected, since each of the terms in this model represents a possible combination of the six PCMH components of interest realized in a practice in our data. We find no significant effect on overall adherence or within any category for patients enrolled in practices with none of the relevant PCMH capabilities (though these are imprecisely estimated). Turning to patients in practices with two or three of the identified PCMH attributes, we find mixed evidence of an effect of specific approaches to PCMH implementation. We see consistent evidence of an increasing PCMH effect on adherence as we move from the four-element practices (especially the Y-Y-Y-Y-N-N configuration) to the 5- and 6-element practices. In the latter two cases, the results are quite similar to the overall PCMH effect results, with significant increases in adherence in all categories except for insulin, and mixed evidence with respect to asthma/COPD medications.

Table 3.5: Impact of PCMH on Medication Adherence (PDC) - Combinations

		Adherence (PDC) by Medication Category							
		Overall	**Diabetes**	**Lipids**	**Insulin**	**Beta Blockers**	**Asthma / COPD**	**Antidep.**	**ACE**
0	**POST x NNNNNN**	-4.608	25.02	-0.562	-5.851	-9.35	5.383	-5.25	30.85
		[19.64]	[21.19]	[19.77]	[4.291]	[18.34]	[5.720]	[8.438]	[19.54]
2	**POST x NNYYNN**	0.61	-6.26	0.268	8.249	3.452**	1.248	0.981	-0.846
		[1.288]	[3.937]	[1.837]	[8.133]	[1.634]	[5.333]	[3.190]	[2.057]
	POST x YYNNNN	1.919*	5.448	0.933	9.427**	-0.295	4.108	3.159	3.428*
		[1.037]	[3.331]	[1.680]	[4.722]	[1.587]	[2.787]	[2.721]	[1.827]
3	**POST x YNYYNN**	0.263	-0.704	0.0779	-10.86*	2.772	-4.078	-3.22	-0.895
		[1.492]	[4.652]	[2.425]	[6.137]	[2.018]	[9.837]	[3.463]	[2.381]
	POST x NNYYNY	1.073	1.578	1.904	3.019	-1.167	0.839	3.323	-0.745
		[1.568]	[3.073]	[1.815]	[7.090]	[1.887]	[6.269]	[3.287]	[2.357]
	POST x YYNNYN	2.235	6.173	1.211	3.797	-1.488	16.84	5.092	-1.746
		[2.798]	[7.822]	[3.924]	[13.79]	[3.684]	[13.87]	[5.869]	[4.705]
	POST x YYYNNN	2.200**	7.275**	0.861	-0.192	6.083***	-0.351	0.499	1.306
		[1.036]	[3.221]	[1.576]	[6.119]	[1.553]	[4.128]	[2.318]	[1.622]
4	**POST x YNYYYN**	0.356	2.891	4.247**	-10.83	1.261	1.07	0.601	-2.66
		[1.393]	[3.487]	[1.894]	[6.843]	[1.877]	[5.262]	[3.229]	[2.413]
	POST x NYYYNY	0.901	-5.859	0.486	15.98**	-5.627**	6.625	2.365	-2.14
		[1.474]	[4.221]	[2.401]	[7.090]	[2.562]	[4.150]	[3.181]	[2.879]
	POST x YYNYNY	2.886	5.371	14.36***	2.778	3.276	-6.534	-13.37**	3.578
		[2.460]	[7.195]	[4.686]	[7.761]	[3.528]	[9.312]	[5.994]	[3.568]
	POST x YYYNYN	3.346***	3.611	0.314	6.568	5.583***	-2.398	4.363*	5.047***
		[1.080]	[2.653]	[1.619]	[6.169]	[1.440]	[4.902]	[2.380]	[1.565]
	POST x YYYYNN	2.148***	3.813***	1.607**	1.28	2.334***	1.855	3.367***	2.961***
		[0.518]	[1.413]	[0.741]	[2.943]	[0.730]	[2.179]	[1.180]	[0.792]
5	**POST x YYYNYY**	4.291***	7.649***	3.647***	-8.194	6.159***	5.327	5.356**	5.215***
		[0.942]	[2.341]	[1.380]	[5.105]	[1.386]	[4.021]	[2.141]	[1.474]
	POST x YYYYNY	3.044***	2.016*	2.766***	-1.559	2.220***	6.659***	4.622***	3.457***
		[0.471]	[1.218]	[0.689]	[2.420]	[0.713]	[1.718]	[1.070]	[0.735]
	POST x YYYYYN	1.987***	3.066***	1.031**	-1.261	2.690***	0.401	3.797***	2.307***
		[0.345]	[0.946]	[0.497]	[1.798]	[0.504]	[1.288]	[0.776]	[0.545]
6	**POST x YYYYYY**	1.804***	3.392***	1.549***	1.03	2.139***	1.771	3.690***	1.982***
		[0.360]	[0.989]	[0.527]	[1.861]	[0.534]	[1.236]	[0.790]	[0.591]
	Obs.	147,496	20,639	68,290	6,148	60,016	20,868	48,534	58,079
	R^2	0.778	0.746	0.739	0.781	0.092	0.842	0.791	0.73

Left column presents number of elements. Coefficients presented are percentage point changes in PDC. Standard errors are clustered at the practice level and appear in brackets. Each model includes the full set of controls listed in the text, but only the DID coefficients are reported. * p<0.1, ** p<0.05, *** p<0.01

Interestingly, adoption of the adherence-related elements is strongly correlated with greater total PCMH recognition scores (see Figure 3.1). Because the analytical approach used here identifies the effect of PCMH adoption using only within-practice variation in patient adherence because of the fixed effects model specification, the effect of post-recognition total PCMH score cannot be separated from the effects of adopting adherence-related PCMH elements. Despite this difficulty, when we consider specific patterns of PCMH implementation (especially those dubbed Y-Y-Y-N-Y-N and Y-Y-Y-Y-N-N here) we observe varying effects, even when holding the number of adherence-related elements constant. This suggests that the choice of specific PCMH elements matters with respect to medication adherence, as opposed to just the overall score.

3.4.3 Parallel Trends Tests

In order for the difference-in-differences method to produce valid estimates of a treatment effect, the assumption of parallel trends must hold – that is, that the same trend would've been observed in the treated and control groups in the absence of the intervention. To check that this condition is satisfied, we use the method from Autor (2003), estimating cluster-specific time effects relative to the year of recognition to account for the multiple time periods covered by our study data as well as the presence of multiple treatment and control groups.

As in the main equation, overall adherence (expressed in PDC) is measured at the individual level while the timing of adoption is measured at the practice level. In each of the six specifications presented, we estimate coefficients for the three pre-periods (t-3, t-2, and t-1), the recognition year (t), and five post-periods (t+1 through t+5), with t-4 as the omitted category. As a result, δ_{Δ} is the coefficient on the Δ^{th} lead or lag across six different specifications: the first specification mimics the PCMH recognition effect regardless of adherence-related elements (as in the top panel of Table 3.4), the second through sixth specification captures the interaction of PCMH recognition with the five groups, $g\epsilon[1,5]$, representing the number of adherence-related elements of PCMH recognition (as in the bottom panel of Table 3.4). Groups

are *g*=1 if 0 or 1 adherence-related elements were adopted, *g*=2 if 2 or 3 adherence-related elements were adopted, *g*=3 if 4 adherence-related elements were adopted, *g*=4 if 5 adherence-related elements were adopted, and *g*=5 if all 6 adherence-related elements were adopted.

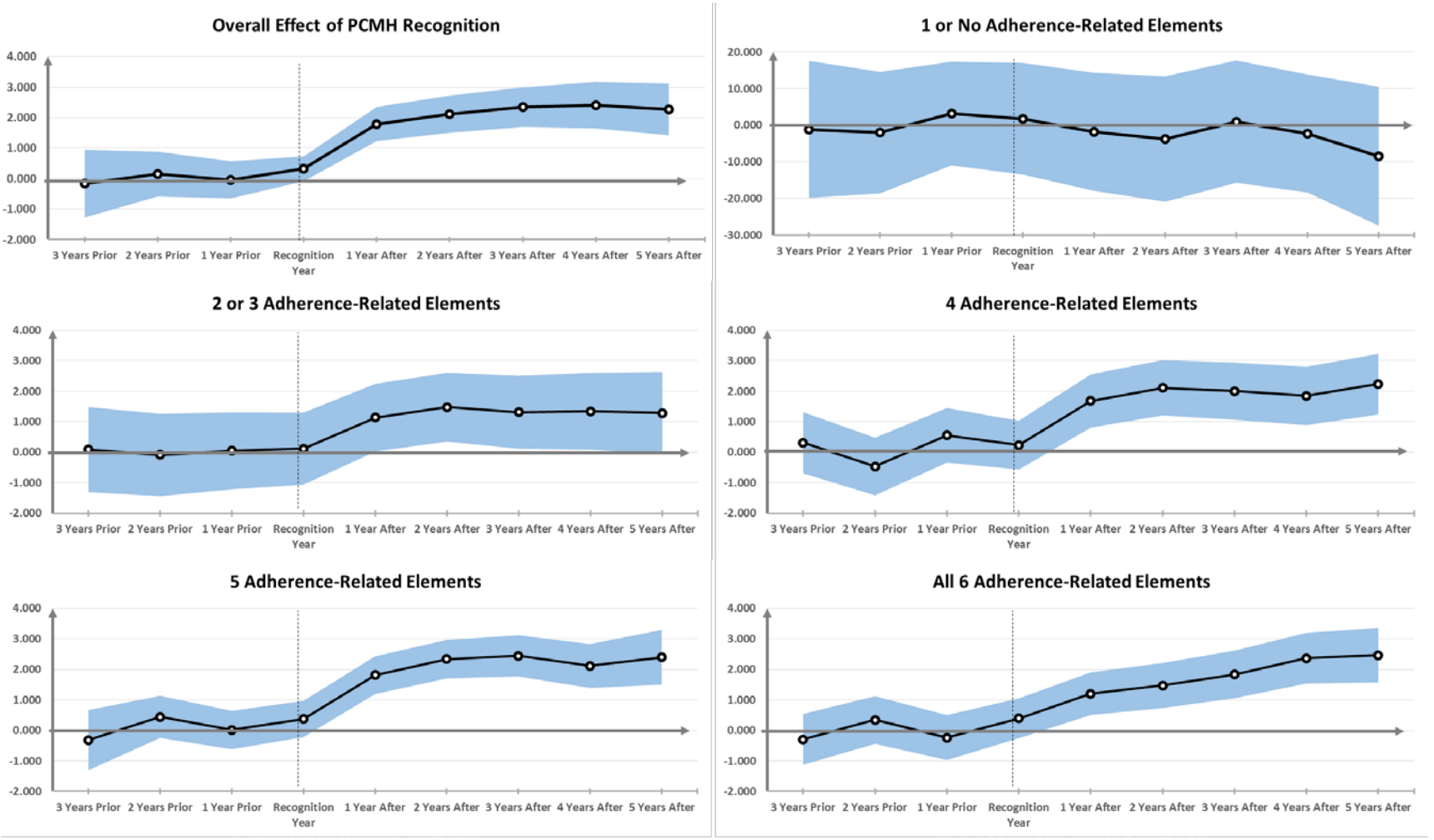
$$Y_{ijt} = \lambda_t + \mu_i + \beta X_{ijt} + \sum_{g=1}^{5} \sum_{\Delta=-m}^{q} \delta_{\Delta} (D_{jt} (\Delta = t - k_j) \times Group_g) + \varepsilon_{ijt} \quad (3.2)$$

Figure 3.2 provides a visual representation of practice-specific trends over time, plotting the coefficient estimates of δ_{Δ} for the specifications above, and the shaded area represents ±1.96 times the standard error of each point estimate. The figure includes six trends: the overall effect in the top-left corner and separate effects for the five groups. A test of the differences-in-differences assumption is $\delta_{\Delta} = 0$ for all $\Delta < 0$. This assumption holds in the overall case as well for the different number of adherence-related elements adopted. With the exception of the case of the adoption of a single or no adherence-related elements (top-right figure), there is an increase in adherence following PCMH recognition. Figure 3.2, indicates that the differences-in-differences strategy seems successful in this context, as the coefficients on the adoption leads are close to zero, showing little evidence of neither an anticipatory response nor implementation preceding recognition. To facilitate comparison of the PMCH effects associated with each number of adopted elements, we have added a single figure that combines the plots from Figure 3.2 in a single graph (with the 0-1 elements results omitted due to scaling) as Appendix Figure C1.

3.5 Discussion

Although evidence has been mixed, there is a substantial body of literature examining the impact of the PCMH model on patient outcomes such as costs and utilization. Considerably less attention has been devoted to the effects of the medical home on medication adherence, despite the importance of this outcome to the management of chronic illness and the emphasis that NCQA has placed on medication-related scoring elements.

Figure 3.2: Time Effects Pre- and Post-NCQA Recognition Year for Overall Adherence by Number of Elements Adopted



Note: Time effects are obtained from regressions of mean change on member fixed-effects, year fixed-effects, and dummies for leads, recognition year, and lags (lead t-4 is the omitted category).

Using detailed, practice-level recognition data not previously available to researchers studying the PCMH model, we are able to explore the impact of specific PCMH capabilities on adherence. Our results suggest that the magnitude and significance of the effect depends on the number of adherence-related recognition components chosen in the implementation of a PCMH. In particular, the effect is not statistically significant for medical homes adopting less than two adherence-related recognition components and is the largest for medical homes with more than three adherence-related components. There is only weak and inconsistent evidence for any of the combinations of 0 to 3 relevant attributes having an effect on adherence. With mixed evidence for 4 attributes, the overall PCMH effect is likely driven by the highest-achieving (5 and 6 elements) practices, which make up the bulk of our sample - nearly 70% of practices and 75% of patients. Our results were also robust to a variety of sample specifications, including analyses of adults with chronic illness and a stable panel of adults with six full years of enrollment in our sample. Moreover, PCMH recognition is based on a mix of practice attributes already in place during the "pre-period" and newly-adopted capabilities at the time of recognition. As a result, our estimation approach should be biased toward zero, providing a lower bound on the PCMH effect on adherence.

Nonetheless, there are two important potential limitations to interpreting the results presented here. First, the patient-specific fixed effects specifications used here control for any time-invariant attributes, and eliminate many potential sources of confounding; however, they cannot control for patient characteristics that vary over time. Second, the dataset used in this study included only patients from a single insurer in a specific region – namely, southeastern Pennsylvania. As a result, some caution should be employed when interpreting these estimates more broadly. Our introduction emphasizes the importance of information asymmetries and uncertain quality to motivate the introduction of the PCMH model. However, whether accreditation *per se* causes quality improvement (or merely signals it) is not definitively answered here. Rather, the most plausible interpretation of our results is that the PCMH recognition process leads practices to adopt improvements to their care management and information technology capabilities which improve adherence.

Given the variety of approaches to capturing adherence (including continuous PDC and the commonly-used 80% threshold used to create a category of "adherent" patients), as well as the focus of a number of interventions on improving outcomes for patients with low baseline adherence (Conn, Ruppara, Enriqueza, & Cooper, 2016), direct comparisons between the current study and others can be difficult. However, a 2012 AHRQ review (Viswanathan et al., 2012) did evaluate the impact of an alternative, policy-based approach to improving adherence: the reduction or elimination of copays for prescription medications. Even though we focus on a different lever, namely reorganization of care delivery, our estimate of a significant, 2.1 percentage-point improvement in continuous PDC falls within the range of estimates provided in the AHRQ review (1.31-6.2 percentage points).

3.6 Conclusion

As the PCMH model continues to receive considerable attention as a vehicle for improving the quality of primary care, understanding whether and how the PCMH model (or particular components of the model) impact medication adherence is of significant policy importance. We find evidence that the PCMH model leads to improved adherence among practices focused on adoption of adherence-related PCMH capabilities. Though a substantial evidence base on the effectiveness of the PCMH model exists, further work is needed to identify whether and how specific elements (or combinations of elements) from the model impact patient care and outcomes.

Chapter 4:

Model Homes: Evaluating Approaches to Patient-Centered Medical Home Implementation

with Guy David, Benjamin Ukert, Abiy Agiro, Sarah Hudson Scholle and Tyler Oberlander

4.1 Introduction

The continuing growth in healthcare costs has pressured insurers to move away from traditional reimbursement systems and typical approaches to care delivery, and instead shift towards models that incentivize improve quality and reduce spending. Arrangements that emphasize process improvements like better care coordination and population health management have the potential to achieve these goals by delivering more evidence-based care, reducing avoidable unnecessary treatments and preventing acute episodes in patients with chronic illnesses. One such model is the Patient Centered Medical Home (PCMH) for primary care improvement. It encourages practices to expand electronic and in-person access to providers, improve coordination and chronic condition management, and utilize information technology to guide and track the care delivered (American Academy of Family Physicians et al., 2007).

The National Committee for Quality Assurance (NCQA) evaluates practices' performance according to six broad categories of improvements ("Standards"), and recognizes qualifying practices at level 1, 2, or 3 (with 3 representing the highest level of performance). PCMH recognition acts as a signal to patients and payers of the high quality of care provided at the practices evaluated by NCQA and, as of 2018, over 13,000 practices have achieved recognition.

The six PCMH standards are made up of 150 specific improvements, or "factors," and practices can achieve recognition by self-selecting a subset of these items to implement. This has led to substantial heterogeneity

in the capabilities among practices with the same PCMH recognition level. A 2014 paper (Tirodkar et al., 2014) described the variation, with the self-explanatory title "there is more than one way to build a medical home." Even among practices recognized as having the same number of PCMH capabilities, a simple "score" to summarize the extent of implementation may conceal important practice differences.

The question remains, however, whether different approaches to PCMH adoption lead to different outcomes. This heterogeneity in the implementation of capabilities also creates a problem from a program evaluation perspective. Among the 11,149 practices recognized under the 2011 NCQA standards, we observe over 8,900 different combinations of the 150 factors which. This incredible diversity in implementation raises questions about the wisdom of trying to identify impact of the medical home model as a single, undifferentiated intervention. Instead, evaluations of the PCMH should follow the course laid out by Ludwig, Kling, & Mullainathan (2011) to seek the specific mechanisms by which the model impacts patient utilization and outcomes. Recent commentary points to this need in studies of the PCMH in particular, highlighting uncertainty in what the PCMH looks like in practice, and calling for research that addresses both how the model is implemented and how medical homes function once put in place (Jackson & Williams, 2015).

Understanding the mechanisms by which the PCMH model impacts patient care becomes particularly important when one considers the persistently high rates of chronic illness among Americans, with more than 40% diagnosed with multiple conditions (Buttorff, Ruder, & Bauman, 2017). Multiple chronic conditions increase the cost and complexity of providing care (Parekh, Goodman, Gordon, & Koh, 2011) and the primary care system in the United States has historically been poorly organized to meet this population's needs (Bodenheimer, Chen, & Bennett, 2009). Efforts to improve care coordination for patients with chronic illness, such as using predictive risk scores to assign patients to care management teams, have shown promise in reducing emergency department (ED) and specialist visits (David, Smith-McLallen, & Ukert, 2019). Substantial progress has been made in other areas of primary care performance as well, such as electronic health record adoption (Office of the National Coordinator for Health

Information Technology, 2016). However, progress has lagged behind in terms of recommended preventive services (Borsky et al., 2018), suggesting a continued need for improving primary care organization, and for a detailed evaluations of how the PCMH model impacts patient care.

Studies of early-adopter PCMH practices found no impact on healthcare spending and limited effects on patient experience (Friedberg et al., 2014; Jackson et al., 2013). More recent work has found reductions in ED utilization (David et al., 2015; Green, Chang, Markovitz, & Paustian, 2018; Rosenthal et al., 2015, 2016) and improved quality of care (Swietek et al., 2018). Of particular note, recent work has shown that practices adopting different capabilities to achieve PCMH adoption experience different changes in utilization outcomes, such as ED visits and specialist care (David, Saynisch, & Smith-McLallen, 2018), and medication adherence (David, Saynisch, Luster, Smith-McLallen, & Chawla, 2018).

This paper estimates the impact of PCMH recognition on patterns of healthcare utilization and expenditures using proprietary claims data from a large commercial insurer from 2006 to 2016. The study combines over five million patient-years spanning 14 states with records on the PCMH capabilities adopted by the roughly 6,000 practices treating them. The analyses utilize detailed practice-level PCMH scoring data not typically available to researchers, which identify the NCQA score in each of the 150 factors. We use hierarchical clustering to develop a typology of PCMH practices based on all 150 NCQA factors. The impact of PCMH recognition on patient utilization and expenditures is then evaluated using a generalized difference-in-differences strategy that takes advantage of variation in the timing of PCMH recognition. First, we assess the impact of PCMH recognition as a whole, treating the medical home as a single, undifferentiated approach to restructuring primary care. Second, we treat membership in each cluster as a different intervention, and assess whether patients in practices taking different approaches to PCMH implementation experience different outcomes.

Our clustering results point to one cluster of practices recognized at relatively low levels of PCMH implementation, and two high-performing clusters. One of the two high-performing clusters distinguishes

itself with respect to an emphasis on adopting electronic communications capabilities. When analyzing the PCMH as a single intervention, our difference-in difference estimates show that PCMH recognition is associated with reduction in emergency department (ED) utilization, as well as reductions in outpatient visits and in utilization of lab and imaging services. Our three cluster difference-in-difference estimates add further clarification, suggesting that while the outpatient visit and service-level results are consistent across clusters, the reduction in ED visits is concentrated in the cluster with expanded electronic communication capabilities.

We contribute to the existing literature in several novel ways. First, to our knowledge, we are the first to utilize a large sample of commercially insured patients receiving care in fourteen states matched to providers who received PCMH recognition. Second, we use proprietary NCQA data that allows us to classify over 6,000 practices based on their capabilities in each of the 150 scoring elements to form distinct practice clusters based on their capabilities. We address identification concerns by studying only practices that apply for PCMH recognition and take advantage of plausibly exogenous variation in the timing of the PCMH application. Third, we are able to observe a large number of healthcare utilization and cost outcome measures to accurately describe the changes in care over time. Fourth, we use data from practices recognized under the revised 2011 NCQA standards, which raised the thresholds for practices to achieve PCMH status.

The rest of the paper is organized as follows. Section 2 provides some background on the PCMH model, and a discussion of previous work to develop typology of medical home practices for use in impact evaluations. Section 3 describes the data used in this study, as well as the analytic approach used to estimate the PCMH effect on patient utilization. Section 4 first describes the results of the practice cluster analysis, and then presents the results of regression analyses of the PCMH effect overall and by cluster. Finally, Section 5 discusses these results in more detail, and Section 6 offers some conclusions.

4.2 Background

Since the publication of the Joint Principles of the Patient-Centered Medical Home in 2007 (American Academy of Family Physicians et al., 2007), NCQA has recognized over 13,000 practices representing 67,000 physicians (NCQA, 2018) for adopting PCMH capabilities. PCMH recognized physicians represent roughly twenty percent of the primary care physician workforce in the US (Cotton, 2018). Though interest in the PCMH and financial support for implementation rapidly increased around the publication of the Joint Principles, the medical home represents a culmination of decades of discussion of how to organize primary care (Kilo & Wasson, 2010).

NCQA began recognizing PCMH practices based on guidelines for the first time in 2008, and this study evaluates practices adopting the 2011 NCQA recognition standards. Relative to the 2008 guidelines, the 2011 standards moved the PCMH health information technology guidelines closer to the Centers for Medicare and Medicaid Services (CMS) Meaningful Use requirements, and added new surveys and procedures aimed at increased patient and family involvement in quality improvement efforts, and increased thresholds for recognition at all three levels (NCQA, 2011). The 2011 criteria are based on six standards divided into 27 elements, of which six are designated as high-importance, "must-pass" elements. The elements are further subdivided into 150 factors. Recognition is granted in one of three levels, with one being the lowest recognition level attained with a score between 35 to 59, two attained with a score of 60 to 84, and three attained with a score of 85 and above out of out of 100 points. These standards and elements are detailed in Table 4.1

In order for practices to achieve recognition by NCQA as a medical home, they are required to submit documentation concerning infrastructure and functions in place. For example, to meet the requirements of Standard 1, Element A ("Access During Office Hours") a practice must document their procedures for scheduling same day appointments, providing clinical advice by phone and other electronic means, and documenting this clinical advice. These self-submitted records are then reviewed by NCQA and a score is issued, with audits of the self-assessments occurring for roughly 5% of practices to ensure compliance. The

Table 4.1: PCMH 2011 Standards and Elements

PCMH 1: Enhance Access and Continuity Element A: Access During Office Hours* Element B: After-Hours Access Element C: Electronic Access Element D: Continuity Element E: Medical Home Responsibilities Element F: Culturally and Linguistically Appropriate Services (CLAS) Element G: The Practice Team	**PCMH 2: Identify and Manage Patient Populations** Element A: Patient Information Element B: Clinical Data Element C: Comprehensive Health Assessment Element D: Use Data for Population Management*
PCMH 3: Plan and Manage Care Element A: Implement Evidence-Based Guidelines Element B: Identify High-Risk Patients Element C: Care Management* Element D: Medication Management Element E: Use Electronic Prescribing	**PCMH 4: Provide Self-Care Support and Community Resources** Element A: Support Self-Care Process* Element B: Provide Referrals to Community Resources
PCMH 5: Track and Coordinate Care Element A: Test Tracking and Follow-Up Element B: Referral Tracking and Follow-Up* Element C: Coordinate With Facilities and Manage Care Transitions	**PCMH 6: Measure and Improve Performance** Element A: Measure Performance Element B: Measure Patient/Family Experience Element C: Implement Continuous Quality Improvement* Element D: Demonstrate Continuous Quality Improvement Element E: Report Performance Element F: Report Data Externally

* indicates each Standard's must-pass Element

practices in this study achieved NCQA recognition between 2011 and 2016, and we use this staggered adoption to study implementation.

Ultimately, the goal of the PCMH model is to reduce cost and improve quality of care for enrolled patients, and the capabilities described in Table 4.1 suggest a number of potential mechanisms. For example, improved population health management, provider adherence to evidence-based care guidelines and increased emphasis on patient self-support education have the potential to reduce exacerbations of chronic illness and avoid the high-costs associated with addressing acute episodes. Additionally, electronic medical records and referral tracking may reduce duplication of services. Expanded communication options like

electronic access and extended practice hours may also allow patients to substitute primary care for expensive alternatives like ED visits. A number of these features have been emphasized in other studies of successful coordinated care programs (Brown, Peikes, Peterson, Schore, & Razafindrakoto, 2012).

The literature analyzing the early impact of the PCMH model found mixed evidence on patient care and outcomes. A systematic review of early studies found only limited effects on patient experience and use of preventive services by patients, with no evidence of reductions in total healthcare expenditures (Jackson et al., 2013). One major concern cited was the lack of consistent definitions of the medical home. These findings were echoed by high-quality studies of early adopter practices which similarly found no significant impacts on cost or utilization (Friedberg et al., 2014) and another study which only found a reduction in emergency department (ED) visits for ambulatory care sensitive diagnoses (Rosenthal et al., 2013). More recently, studies of the PCMH model has found that adoption of the model leads to reduced utilization of high-intensity medical services like ED and hospital visits among patients with chronic illnesses (David et al., 2015; Green et al., 2018), and improves quality of care for patients with multiple chronic conditions (Swietek et al., 2018). Further studies have found reductions in ED utilization and costs looking at a wider array of patient populations (Rosenthal et al., 2015, 2016). However, conflicting evidence continues to be generated about the PCMH model's effectiveness, including a 2017 meta-analysis which found no effect on primary care, ED visits or inpatient stays across a range of PCMH pilot programs (Sinaiko et al., 2017).

One major limitation of these studies has been the limited ability to pair detailed data on the specific approach to implementation practices have taken. Exceptions to this include the use of surveys to establish what PCMH practices and infrastructure were adopted (Friedberg et al., 2014; Poznyak, Peikes, Wakar, Brown, & Reid, 2018) and reliance on sources of data other than the NCQA recognition scoring guidelines, including the Patient Aligned Care Teams (PACT) initiative data from the Veterans Health Administration (Rosland et al., 2018). This study represents another effort towards studying both the implementation of the PCMH model and its effectiveness, as suggested by (Jackson & Williams, 2015), using detailed data from

the NCQA's own PCMH recognition database to identify which capabilities practices had in place at the time of recognition.

Additionally, the only other papers to date that link detailed NCQA capability data to patient outcomes are David, Saynisch, & Smith-McLallen (2018) and David, Saynisch, Luster, et al. (2018). This study builds on their methodology of hierarchical clustering, but advances the knowledge in this area significantly. As described in the introduction, prior research which did use the NCQA recognition data was limited by, among other things, a narrow geographic focus and analyzed only the impact of the initial 2008 NCQA recognition standards.

4.3 Methods

4.3.1 Data Sources

We combine two sources of data to assess the impact of the PCMH model on patterns of healthcare utilization: NCQA's detailed records of practice-level PCMH capabilities; and medical claims data from HealthCore Integrated Research Database (HIRD), which contains medical and pharmacy administrative claims and health plan eligibility data from Anthem's Blue Cross/Blue Shield (BCBS) commercial health plans. Our practice-level data are drawn from NCQA recognition records, and provide detailed documentation on the specific capabilities each practice had in place at the time of recognition at the factor level. The availability of these records for the complete set of practices recognized under NCQA's 2011 PCMH Standards represents a significant advance in the ability to evaluate the effect of PCMH adoption on patient care.

The data on practice capabilities were then merged to the HIRD, which contains medical and pharmacy administrative claims and health plan eligibility data from Anthem's fourteen Blue Cross/Blue Shield (BCBS) commercial health plans. Records were available for covered patients from 2006 to 2016. Our patient control variables include socio-demographic variables (age and gender) health profile (Deyo-Charlson comorbidity index (Deyo, Cherkin, & Ciol, 1992)) and indicators for whether the patient had a

diagnosis of chronic obstructive pulmonary disease (COPD), congestive heart failure (CHF), diabetes or any malignancy in the last year.

Our utilization outcomes were captured with two types of measures: binary indicators capturing whether there was any utilization of a given service in a year (extensive margin); and a second, continuous measure of the number of utilization conditional on a patient-year observation featuring at least one visit (intensive margin). Our utilization measures span the universe of healthcare system interactions and includes utilization of inpatient admissions, ED visits, preventable ED visits[37], general physician visits[38], specialist visits, the use of laboratory services, and the use of diagnostic imaging services. We also inflation adjusted expenditures for each of these categories, and also evaluated total healthcare expenditures. We calculate healthcare expenditures as the sum of amounts paid by both the insurer and patient (co-pays and co-insurance).

We made several restrictions to our sample to identify how PCMH recognition affects utilization and expenditures. First, we identify all patients with at least two claims for evaluation and management visits with a given primary care provider (PCP). The requirement of two visits, with one before and one after PCMH recognition, is necessary to allow for the use of patient fixed effects in the regression analyses. Second, we retain only patients with a full year of continuous enrollment before and after the date of PCMH recognition to ensure that we have adequate data to estimate the effect of PCMH recognition using patient fixed effects. Third, in order to cleanly attribute patients to a provider (and PCMH practice) of record, we exclude patients matched to multiple physician NPIs.[39] Fourth, we drop the year of recognition from our main analysis to address disruption that may result from introduction and improvement of capabilities may

[37] Visits defined here as preventable are based on work published by the Commonwealth Fund (Billings, Parikh, & Mijanovich, 2000). A "preventable" visit is one with a greater cumulative probability of falling in the non-emergent, primary care treatable, or avoidable categories.
[38] We define general physician visits as those identified on the claim with a type of service for Evaluation and Management performed by a primary care physician (PCP). PCPs are defined as physicians with one of the following specialties: general physician, family physician, geriatrics, and internal medicine.
[39] We match the physician NPIs from those claims to data from NCQA on physician-practice correspondence to link patients to a practice and identify the year in which the PCMH recognition occurred.

occur over the course of the year. Finally, we exclude patients younger than 18, who were likely treated in different (pediatric) practice environments.

These restrictions limit the impact of issues with patient attribution during years that were only partially visible or treated. Since the PCMH model is especially promising for members who benefit from continuity of care and who have chronic conditions, focusing on patients with an existing history with the provider is beneficial for an analysis of the PCMH model. Additionally, dropping the year of PCMH recognition from the analysis accounts for potential challenges in implementation that may lead to short-term disruptions in practice (Berenson et al., 2008; Harbrecht & Latts, 2012) and is consistent with other health policy evaluations (Chandra et al., 2011; Joynt et al., 2013).

Table 4.2 presents summary statistics for patient characteristics, healthcare utilization and healthcare expenditures for the full sample and for the three clusters identified from the hierarchical cluster modeling described in the following section. Cluster 1 stands out from the other two clusters as being slightly older and sicker, both in terms of comorbidity scores and the rates of specific comorbidities. These differences are also reflected in terms of typical utilization, with higher average annual spending and a greater probability of using high-intensity medical services such as inpatient admissions or ED visits.

4.3.2 Empirical Approach

We analyze the data on PCMH practices and patient utilization in two steps. First, we employ hierarchical agglomerative clustering to group the practices into maximally similar categories based on their PCMH capabilities. Similar techniques have been used elsewhere in the health services literature to develop a typology of primary care practices (David, Saynisch, & Smith-McLallen, 2018; Shortell et al., 2014). Hierarchical clustering begins with N clusters which each contain one observation, which are combined sequentially until a single cluster of N observations remains. At each step, a proximity matrix that contains the distances between each possible pair of clusters is calculated. Each factor has binary scoring (that is,

Table 4.2: Summary of Patient Characteristics

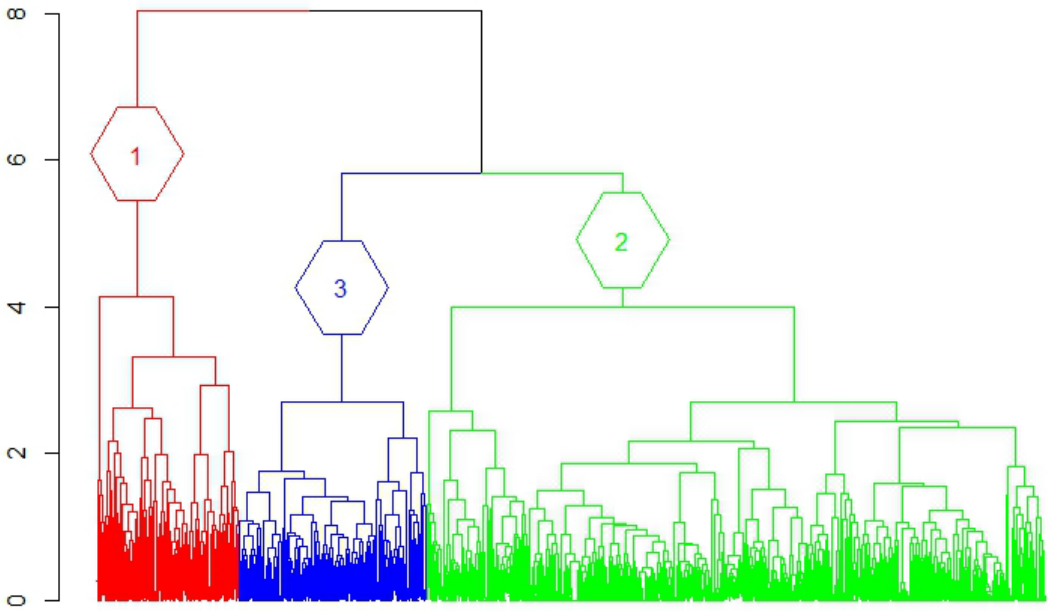
	Total		Cluster 1		Cluster 2		Cluster 3	
	N=5,313,917		N=427,174		N=4,195,442		N=691,301	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Patient Characteristics								
Age	52.34	16.13	54.58	16.37	52.13	16.07	52.25	16.25
Gender (Female)	0.55	0.50	0.55	0.50	0.54	0.50	0.55	0.50
Deyo-Charlson comorbidity score	0.59	1.25	0.68	1.34	0.57	1.23	0.60	1.25
Congestive heart failure	0.02	0.14	0.03	0.16	0.02	0.14	0.02	0.14
Chronic obstructive pulmonary disease	0.10	0.30	0.12	0.33	0.10	0.30	0.10	0.30
Malignancy	0.05	0.23	0.06	0.23	0.05	0.23	0.05	0.22
Diabetes complications	0.03	0.16	0.03	0.18	0.03	0.16	0.03	0.17
Percent with Any Use								
Inpatient visit	0.07	0.26	0.08	0.28	0.07	0.26	0.07	0.26
ED visit	0.13	0.34	0.15	0.35	0.13	0.34	0.14	0.35
General Physician [1]	0.74	0.44	0.72	0.45	0.74	0.44	0.74	0.44
Specialist Visit [2]	0.66	0.47	0.67	0.47	0.66	0.47	0.65	0.48
Imaging Service	0.56	0.50	0.58	0.49	0.56	0.50	0.55	0.50
Lab Service	0.76	0.43	0.76	0.42	0.76	0.43	0.76	0.43
Mean Use, Conditional on >0								
Inpatient visit	1.33	0.90	1.35	0.98	1.33	0.89	1.33	0.92
ED visit	1.32	0.95	1.35	1.01	1.31	0.93	1.35	1.07
General Physician [1]	2.83	2.33	2.97	2.46	2.80	2.30	2.95	2.44
Specialist Visit [2]	4.09	4.26	4.15	4.20	4.06	4.22	4.20	4.47
Imaging Service	3.79	3.68	3.94	3.83	3.77	3.67	3.78	3.66
Lab Service	4.65	5.14	4.62	4.90	4.66	5.16	4.63	5.16
Total costs	$6,358	$22,633	$6,902	$22,595	$6,284	$22,573	$6,470	$23,006

[1]Defined as an evaluation and management visit to a general practice, family practice, internal medicine, or geriatric medicine

[2]Defined as an evaluation and management to physicians other than general practice, family practice, internal medicine or geriatric medicine

the capability was present or absent) so the proximity is measured using the Jaccard distance.[40] Once identified, the pair with the minimum distance is merged. After merging, the matrix is recalculated to reflect the distances between the remaining clusters (Aggarwal & Reddy, 2014). In this analysis, Ward's criterion is used, which seeks to minimize the increase in within-cluster variance as clusters are merged (Ward, 1963). This process can be summarized in a dendrogram, in which each node, or "leaf," represents a single practice at the beginning of the process, and the terminal node, or "trunk," displays the resulting single cluster of all practices. The change in height at each step in the dendrogram represents the relative increase in within-cluster distances after each step. The dendrogram generated in this clustering procedure appears in Figure 4.1.

Figure 4.1: Clustering Dendrogram



[40] The Jaccard similarity coefficient is the ratio of the number of factors shared by both practices (the intersection) divided by the total number of factors present in either practice (the union). Similar clusters and patterns of cluster emphasis are achieved using alternative measures, like the simple matching coefficient (SMC).

Starting from the terminal node and moving toward the leaves, the within-cluster variation decreases substantially moving from one to two clusters, and to a lesser extent, from two to three. After this step, the differences in within-cluster variation are much smaller[41]; as a consequence, a three-cluster solution was chosen.

We then use a generalized difference-in-differences framework to estimate the effect of PCMH recognition on patient utilization, identifying the impact using variation in the timing of adoption. We proceed using two sets of regression specifications. First, we use the conventional approach, which implicitly treats the PCMH as a single, undifferentiated intervention (Equation 1). To complement this method, we then treat the three clusters as separate treatments, each with their own difference-in-differences term (Equation 2). By comparing these results, we can assess whether and to what extent the clusters vary in their impact on patterns of patient utilization, and to what extent treating the PCMH as a homogenous approach to practice restructuring may obscure important variation.

$$Y_{ijt} = \lambda_t + \mu_{i/j} + \beta X_{it} + \delta\, Post_{it} + \varepsilon_{ijt} \quad (4.1)$$

$$Y_{ijt} = \lambda_t + \mu_{i/j} + \beta X_{it} + \sum_{g=1}^{G} \delta_g \left(Post\ x\ Cluster_g\right)_{it} + \varepsilon_{ijt} \quad (4.2)$$

In both specifications, the term Y_{ijt} represents one of the utilization or expenditure outcomes listed in the previous section for individual *i* treated by physician *j* in year *t*. Year fixed effects, expressed with λ_t control for secular trends in clinical practice over the study period. Additionally, we include either patient or practice fixed effects ($\mu_{i/j}$) in each model to control for unobserved, time-invariant features of practice or individual patients. *X* is a vector of time-variant patient controls. Finally, *Post* and *Post x Cluster* are the difference-in-difference terms, which are equal to one if a patient-year is associated with a practice that has

[41] In the dendrogram presented in Figure 4.1, the first split (producing two clusters) occurs at a height of 8, the second at roughly 5.8, and the third at 4.2. The next split occurs at a height of 4, a much smaller gap.

achieved PCMH recognition, and zero otherwise. The utilization regressions with binary dependent variables are estimated as linear probability models.

Using the clusters as indicators of separate treatments has several advantages over alternative approaches. One such approach would be to simply include the PCMH capabilities detailed in the recognition data as regression controls. Once even first-order interactions between features are considered, however, the number of controls quickly outpaces the sample of practices. This design would also be affected by the issue that efforts to estimate the marginal contribution of any one PCMH feature to patterns of utilization an expenditure may not be interpretable, given the high level of complementarity between features. In light of the limitations facing alternative methods, clustering has substantial appeal as a pre-processing step. It relies on actual patterns of PCMH adoption and therefore avoids any possibility of estimating the effect of groups of PCMH capabilities that are not typically implemented together, and relies on a small number of terms being added to any regression. In other words, clustering is a strong fit for our goal of identifying and evaluating approaches to PCMH implementation actually being used in primary care practice, rather than estimating what the impact of hypothetical alternative approaches might be.

Finally, to address the potential of self-selection into the PCMH recognition process to bias our estimates, our sample is limited to practices which eventually received recognition as medical homes during the 2006 to 2016 study period. This approach addresses a number of issues that would otherwise impact and potentially bias our results. Primarily, limiting our analyses to practices which eventually achieved recognition reduces the impact of self-selection into treatment will bias our results. Previous research has indicated that practices that never adopted the PCMH during this window were systematically different from those that did (David, Saynisch, & Smith-McLallen, 2018). As a result, comparing these two groups may erroneously attribute the impact of factors like differences in administrative capacity to the effect of PCMH adoption. Moreover, the analyses that follow use patient and practice fixed effects to control for unobserved, time-invariant features of the units studied. As a consequence, practices which never switch into treatment (and patients into those practices) will have no impact on the regression estimates of the

PCMH effect. Finally, the detailed practice capability data is only available for practices that eventually underwent the recognition process.

4.4 Results

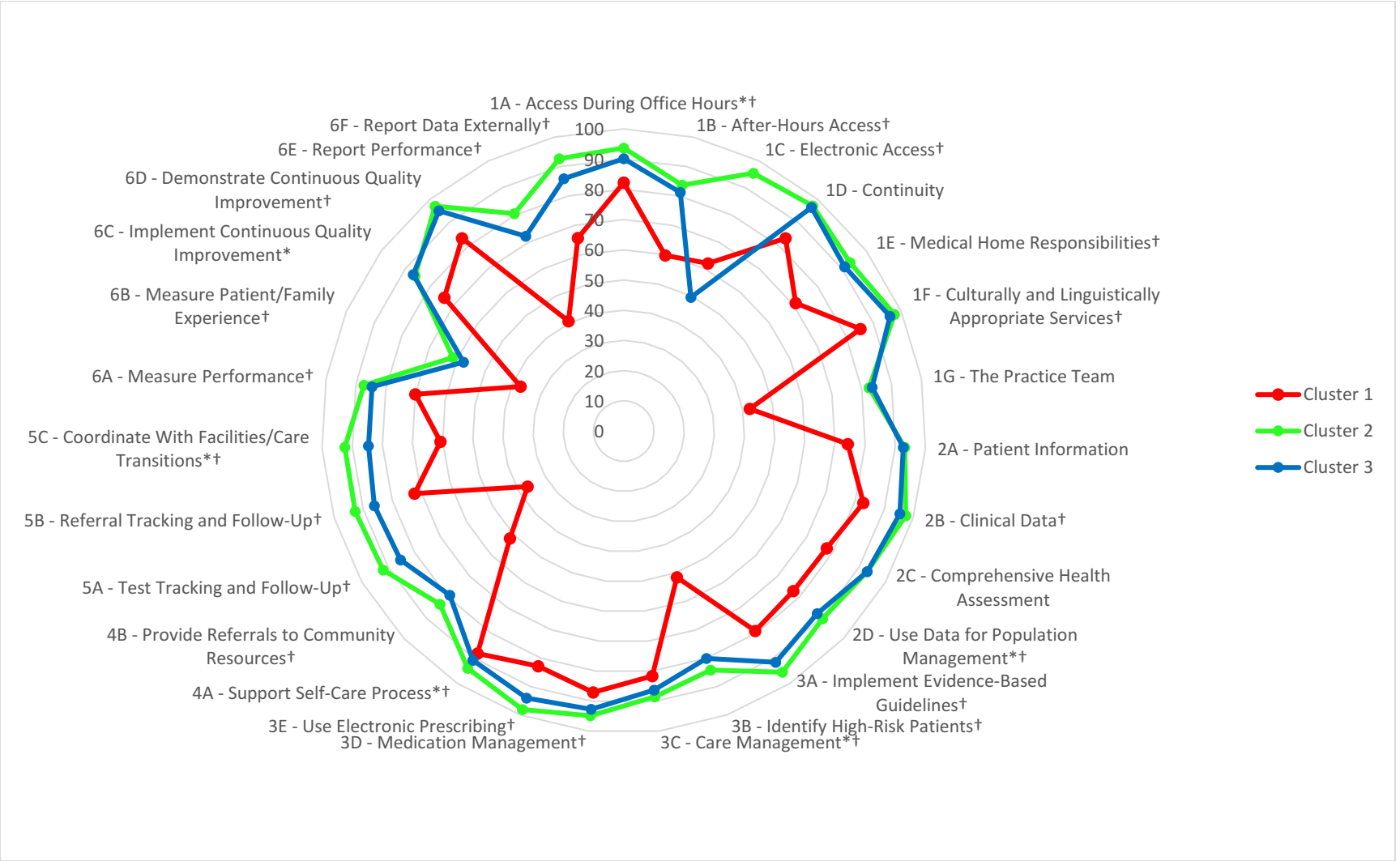
4.4.1 Hierarchical Clustering Results

To concisely summarize the features of the three clusters produced by the hierarchical clustering algorithm, we aggregated scores for each of the 27 PCMH elements by cluster. The relative performance of each cluster is presented as a radar plot in Figure 4.2, displaying the score in each element as a percent of the maximum achievable. The average performance in a given cluster in each element is expressed as a node along a radius, with an average score of 100 percent appearing at the perimeter of the figure and an average score of zero represented at the center. In other words, a higher element score, and therefore capability, is displayed with a greater distance from the center of the plot. A visual inspection of the plot shows that Cluster 1 is consistently lower performing than Cluster 2 and 3. While Cluster 2 outperforms Cluster 3 in the majority of the elements, the differences is most noticeable in the "electronic access" element. The detailed percent scores by element appear in the first 3 columns of Appendix Table D1, and a more detailed set of radar plots at the factor level are presented in Appendix Figure D1.

The full, factor-level plots do point to some qualitatively large gaps between Clusters 2 and 3 in factors outside of Element 1C; however, these generally refer to electronic communication capabilities that were simply grouped under another heading. For example, one such gap is visible for Element 1A Factor 3, which corresponds to "providing timely clinical advice by secure electronic messages during office hours." Thus, practices' implementation of electronic communications infrastructure appears to be the main difference between these clusters, regardless of the level of PCMH data used to describe them.

We performed a series of one-way ANOVA tests to assess whether the elements within each of the three clusters were statistically different. Given that we estimate 27 one-way ANOVA test in total (one for each element), we used the conservative Bonferroni correction to address the multiple hypothesis testing and set

Figure 4.2: Cluster Performance by PCMH Element



* indicates each Standard's must-pass Element

† indicates significance differences between Clusters 2 and 3 at the $\alpha = 0.05/27 = 0.0019$ level

the significance thresholds to 0.0019 (that is, $\alpha = 0.05/27$). Even at the Bonferroni p-level threshold the clusters were significantly different in each element, including the six "must-pass" elements.

We conducted a second series of ANOVA tests testing for element differences in the high performing clusters with the same conservative Bonferroni p-value correction. This latter analysis pointed to significant differences for all but five of the elements. However, many of the differences are qualitatively small and likely of limited clinical significance. For example, with respect to all elements in Standard 1 (1A-1G), the average difference between clusters 2 and 3 is only 3.5 percentage points. One element, however, stands out with a stark contrast between the two high-performing clusters: the difference between Cluster 2 and Cluster 3 in Element 1C (Electronic Access) is 46 percentage points. Columns 4 and 5 in Appendix Table D1 display all ANOVA test p-values.

4.4.2 *Regression Analysis*

We now turn to estimating the impact of PCMH adoption on patient utilization and expenditures. As described in the previous section, we first estimate an overall effect of PCMH adoption in our sample, which treats PCMH adoption as a single, undifferentiated treatment. Then, we re-estimate models using three separate treatment indicators, one for each cluster that resulted from the hierarchical clustering algorithm. Estimates using practice and patient fixed effects are presented side-by-side, but generally produce consistent results. In the tables that follow, we present the estimates for the PCMH effect, or δ from equations (4.1) and (4.2), which are interpretable as percentage point changes in the specified utilization outcome, changes in the number of visits, or percent changes in the case of log-transformed costs. Each estimate is followed by standard errors (clustered either at the practice- or patient-level) in brackets and, for statistically significant results, a translation of percentage point estimates into percent changes as compared to baseline rates of the specified outcome. The results are organized as follows:

Table 4.3 presents the overall PCMH results (Equation 4.1) with respect to extensive margin effects and log-transformed total healthcare costs; Table 4.4 presents the by-cluster results (Equation 4.2) for the

extensive margin effects and log costs; and Table 4.5 presents the by-cluster results for the intensive margin effects. Supplemental tables showing the overall PCMH intensive margin effects and effects on log-transformed costs by category appear in the appendix.[42]

Table 4.3 presents the results for the overall effect of PCMH adoption on care utilization, with respect to extensive margins. We find no effect of PCMH adoption on inpatient utilization, even though our models are precisely estimated. For all other categories, we find significant reductions in utilization, regardless of whether patient or practice fixed effects are included (though the point estimates using patient fixed effects are consistently smaller). Specifically, PCMH adoption lead to a 1.9-2.5% reduction in ED utilization, depending on the specification. Use of imaging fell by 1.1-1.4% and utilization of laboratory services fell by 1.5-2.3%. The bottom-right panel of Table 4.3 presents the results for the effect on total healthcare expenditures, and point to an 8.3% reduction in costs. Consistent with the underlying expected mechanisms, specialist utilization fell; though surprisingly, general physician use fell as well. This finding may suggest that the general PCMH model improves care coordination without any adverse consequences in terms of hospitalization or ED visits.

Table 4.4 presents cluster specific results for the PCMH effect on patient utilization on the extensive margin – that is, changes in the probability of any visit or utilization of a service. We find a number of outcomes where analyzing the PCMH as a single intervention obscures variation. The precisely-estimated overall zero effect on inpatient admissions concealed a small *increase* in hospitalizations among patients in Cluster 1, the relatively low-performing practices. Moreover, the reductions in ED utilization were concentrated in Cluster 2, which was the high-performing cluster that emphasized electronic communications, and the effect was larger than in the overall estimates (2.3-3.1% reductions). The reductions in general physician and specialist utilization were largest in Cluster 1 as well, with the smallest reductions in Cluster 3. That pattern is reversed for use of imaging and lab services, for which Cluster 3 practices experienced the largest

[42] Across the individual categories of spending, the patterns of sign and significance of the changes mirrored those for the extensive margin results presented in Table 4.5.

Table 4.3: Overall PCMH Adoption Effect Estimates, Utilization

	Patient FE	Practice FE
Inpatient visit	0.0003	0.0003
	(0.0007)	(0.0007)
	0.47%	0.37%
ED visit	-0.0024***	-0.0033***
	(0.0009)	(0.0011)
	-1.85%	-2.54%
General physician visit	-0.0141***	-0.0225***
	(0.0011)	(0.0038)
	-1.91%	-3.04%
Specialty visit	-0.0069***	-0.0100***
	(0.0011)	(0.00234)
	-1.05%	-1.52%
Imaging service	-0.0064***	-0.0080***
	(0.0012)	(0.0020)
	-1.14%	-1.43%
Lab service	-0.0117***	-0.0171***
	(0.0010)	(0.0028)
	-1.54%	-2.25%
Log Total Cost	-0.0833***	-0.111***
	(0.0050)	(0.0177)

Reported coefficients are for overall PCMH indicator. Standard errors appear in parentheses, and percent change is calculated by dividing the percentage point coefficient estimate by average outcome rate. Results exclude observations with partial patient-year or which occurred during the PCMH recognition year. Regression model also adjusted for year (2006-2016), age, age-squared, gender, comorbidity index, COPD, CHF, malignancy, and diabetes.

*** p<0.01, ** p<0.05, * p<0.1

reductions. With respect to changes in spending, patients in Cluster 1 experienced an average reduction of 8.7% in annual costs, whereas patients in Clusters 2 and 3 both saw reductions of 8.3%. These heterogeneous findings suggest that the subset of capabilities practices adopt plays an important role in how care is organized and utilized.

Table 4.5 presents estimates of the cluster specific PCMH effect on the intensive margin, or the number of visits for patients with non-zero utilization in a given year. These results can provide additional insights into how utilization changed for patients with greater severity of illness. We find *greater* numbers of ED

Table 4.4: PCMH Adoption Effect Estimates by Cluster, Extensive Margin

	Inpatient visit		ED visit		General physician visit	
	Patient FE	**Practice FE**	**Patient FE**	**Practice FE**	**Patient FE**	**Practice FE**
Cluster 1 x Post	0.0020*	0.0016	-0.0002	-0.00003	-0.0193***	-0.0263***
	(0.0012)	(0.0013)	(0.0015)	(0.0021)	(0.0019)	(0.0082)
	2.50%				-2.68%	-3.65%
Cluster 2 x Post	0.0002	0.0001	-0.0030***	-0.0040***	-0.0146***	-0.0228***
	(0.0007)	(0.0007)	(0.0009)	(0.0011)	(0.0011)	(0.0037)
			-2.31%	3.08%	-1.97%	-3.08%
Cluster 3 x Post	0.0002	0.0002	0.0002	-0.001	-0.0079***	-0.0186**
	(0.0010)	(0.0011)	(0.0013)	(0.0021)	(0.0016)	(0.0074)
					-1.07%	-2.51%
	Specialist visit		**Imaging service**		**Lab service**	
	Patient FE	**Practice FE**	**Patient FE**	**Practice FE**	**Patient FE**	**Practice FE**
Cluster 1 x Post	-0.0072***	-0.0151***	-0.0058***	-0.0086**	-0.0069***	-0.0101*
	(0.0019)	(0.0047)	(0.0020)	(0.0041)	(0.0018)	(0.0060)
	-1.07%	-2.25%	-1.00%	-1.48%	-0.91%	-1.33%
Cluster 2 x Post	-0.0076***	-0.0096***	-0.0062***	-0.0074***	-0.0121***	-0.0175***
	(0.0011)	(0.0022)	(0.0012)	(0.0019)	(0.0011)	(0.0026)
	-1.15%	-1.45%	-1.11%	-1.32%	-1.59%	-2.30%
Cluster 3 x Post	-0.0020	-0.0095**	-0.0083***	-0.0117***	-0.0121***	-0.0187***
	(0.0017)	(0.0046)	(0.0017)	(0.0035)	(0.0016)	(0.0051)
		-1.46%	-1.51%	-2.13%	-1.59%	-2.46%
			Log Total Costs			
			Patient FE	**Practice FE**		
Cluster 1 x Post			-0.0868***	-0.131***		
			(0.0087)	(0.0353)		
Cluster 2 x Post			-0.0829***	-0.106***		
			(0.0051)	(0.0149)		
Cluster 3 x Post			-0.0836***	-0.129***		
			(0.0077)	(0.0353)		

Reported coefficients are for Cluster x Post indicators. Standard errors appear in parentheses, and percent change is calculated by dividing the percentage point coefficient estimate by average outcome rate. Results exclude observations with partial patient-year or which occurred during the PCMH recognition year. Regression model also adjusted for year (2006-2016), age, age-squared, gender, comorbidity index, COPD, CHF, malignancy, and diabetes.

*** p<0.01, ** p<0.05, * p<0.1

visits for patients with at least one visit in both Cluster 1 (1.7-3.5% increase) and Cluster 2 (0.9-2.0% increase) – results that hold, even when the dependent variable is limited to potentially preventable ED visits. Additionally, we find no change in the number of general physician visits for Clusters 1 and 3, but

do find a significant reduction in visits for patients in Cluster 2. The reductions in the probability of any specialist utilization shown in Table 4.4 are echoed in Table 4.5, with each cluster showing reduced utilization on the intensive margin as well. The largest effects were observed for practices in Cluster 1, with a 1.8% reduction in the patient fixed effects specification. Turning to use of imaging and labs, we find no significant change in the number of imaging services used in Clusters 2 and 3, though we do find a reduction in Cluster 1. With respect to labs, we find that the patients with any utilization actually had substantially greater consumption after PCMH adoption (7.0-8.3%), whereas patients in Clusters 2 and 3 had small reductions in conditional use.

4.5 Discussion

Despite mixed evidence from early evaluations of the PCMH model, implementation has continued apace, reaching roughly twenty percent of the US primary care physician workforce. A vast patient population has experienced major changes in their care, and practitioners have faced potentially high cost of adopting and maintaining medical home infrastructure ($30,991 and $147,573 on average, respectively) (Martsolf, Kandrack, Gabbay, & Friedberg, 2016). As a consequence, understanding how practices are approaching medical home recognition – and what approaches work – is essential. The clustering approach presented in this study points to three groups of practices in the wave of recognitions using the 2011 NCQA guidelines: one cluster made of practices at a relatively low level of PCMH performance, and two high-performing clusters. In the higher performing pair, Cluster 2 consistently achieved higher levels of PCMH adoption than Cluster 3, with an especially large gap in terms of adoption of electronic communications.

This "nested" comparison of the two high-performing clusters stands in contrast to earlier work on developing a typology of medical home implementation (David, Saynisch, & Smith-McLallen, 2018). In this prior research, the clustering pointed to starkly different approaches to implementation among the highest-performing practices: in that case, one cluster focusing on "physician-facing" features like decision-supports; and another emphasizing "patient-facing" capabilities like population health management. One

Table 4.5: PCMH Adoption Effect Estimates by Cluster, Intensive Margin

	Inpatient visit		ED visit		General physician visit	
	Patient FE	**Practice FE**	**Patient FE**	**Practice FE**	**Patient FE**	**Practice FE**
Cluster 1 x Post	0.0176	0.009	0.0475***	0.0233*	-0.0013	-0.0312
	(0.0261)	(0.0148)	(0.0167)	(0.0124)	(0.0118)	(0.0237)
			3.52%	1.73%		
Cluster 2 x Post	0.0149	0.0134	0.0257**	0.0113*	-0.0406***	-0.0440***
	(0.0159)	(0.00936)	(0.0112)	(0.00674)	(0.00621)	(0.0122)
			1.96%	0.86%	-1.45%	-1.57%
Cluster 3 x Post	0.0123	0.0110	0.0244	0.0112	-0.00602	-0.0143
	(0.0222)	(0.0127)	(0.0160)	(0.0113)	(0.00971)	(0.0198)

	Specialist visit		Imaging service		Lab service	
	Patient FE	**Practice FE**	**Patient FE**	**Practice FE**	**Patient FE**	**Practice FE**
Cluster 1 x Post	-0.0725***	-0.0672*	-0.0455**	-0.0768**	0.323***	0.385***
	(0.0213)	(0.0358)	(0.0228)	(0.0336)	(0.0250)	(0.0872)
	-1.75%	-1.62%	-1.15%	-1.95%	6.99%	8.33%
Cluster 2 x Post	-0.0547***	-0.0397**	-0.0163	-0.0187	-0.0346**	-0.0429
	(0.0122)	(0.0170)	(0.0144)	(0.0207)	(0.0144)	(0.0364)
	-1.35%	-0.98%			-0.74%	
Cluster 3 x Post	-0.0386**	-0.0574**	-0.0011	-0.0151	-0.0410*	-0.0249
	(0.0189)	(0.0254)	(0.0201)	(0.0272)	(0.0218)	(0.0531)
	-0.92%	-1.37%			-0.89%	

Reported coefficients are for Cluster x Post indicators. Standard errors appear in parentheses, and percent change is calculated by dividing the percentage point coefficient estimate by average outcome rate. Results exclude observations with partial patient-year or which occurred during the PCMH recognition year. Regression model also adjusted for year (2006-2016), age, age-squared, gender, comorbidity index, COPD, CHF, malignancy, and diabetes.
*** p<0.01, ** p<0.05, * p<0.1

possibility is that this is simply a consequence of the larger, more geographically diverse sample of practices included in this study as compared to the 104 included in the previous work. Alternatively, the prior analysis focused on early adopters using the 2008 NCQA guidelines, whereas this study employs a set of revised standards published in 2011. The combination of streamlined recognition standards and a greater accumulation of experience with the medical home model and recognition process among physicians may have led to a more standardized approach, explaining the "nested" appearance of the clusters presented in

the previous section. More work is needed to unpack what features of practices, health systems and patient populations have driven differing approaches to medical home adoption.

Despite these contrasts, the difference-in-differences analysis of the clusters provides an opportunity to study whether and how different approaches to implementation impact patient outcomes. We find no reduction in inpatient admissions following medical home adoption (and find a marginally-significant increase in Cluster 1, though only in the patient fixed effects specification). Similar to a number of prior papers, we find significant reductions in ED visits in PCMH practices. Turning to the results by cluster, we see that this was driven entirely by Cluster 2, the high-performing practices which emphasized electronic communications. One possible mechanism for this change is that expanding electronic access, particularly after hours, might offer a substitute for low-acuity ED visits. This explanation is supported by the finding that Cluster 2 practices also experienced a significant reduction in potentially preventable ED visits.

Consistent with expectations that the PCMH model would reduce utilization of higher-intensity services, we observe a reduction in specialist visits in the overall analyses presented in Table 4.3, with the effect concentrated among patients attributed to Cluster 1 and 2 practices. Contrary to these expectations, however, we see large, significant reductions in general physician visits overall. Results presented in Table 4.5 concerning the PCMH effects on intensive margins do not suggest that this is merely a consequence of physician time being reallocated to higher-severity patients in these practices. Somewhat reassuringly, we do not see increases in ED visits or inpatient admissions, but without detailed data on health (as opposed to utilization) outcomes, it is difficult to draw strong conclusions about how this reduction in the probability of a visit impacts patient welfare. Of particular note is the evidence on reductions in overall spending following PCMH recognition. All three clusters experienced reductions in spending of over eight percent in the analyses using patient fixed effects. This effect appears to be driven by reductions in outpatient spending, both in terms of physician fees and the use of labs and imaging. While prior research has pointed to high fixed and variable costs associated with PCMH recognition, this work also points to a wide range of costs, with over five-fold variation in per-physician spending on PCMH maintenance (Martsolf et al.,

2016). The large reductions in total costs documented here suggest that payers may benefit from continued support for PCMH implementation, and call for further research into how to maintain PCMH infrastructure in the most cost-effective manner.

This paper has a number of limitations. While the sample size of patients and practices is large and geographically diverse, it includes claims data from only a single payer and cannot address how the impact of PCMH implementation may vary with the proportion of a patient panel affected. Additionally, this study focuses on healthcare utilization rather than health outcomes per se. As a consequence, any positive impacts with regard to reduced utilization or spending must come with the caveat that there may be unobserved, negative consequences on some dimension of patient health. Finally, with respect to the estimation approach used, the two sets of specifications presented use patient and practice fixed effects. Because they are able to control for aspects of patient health status that might otherwise bias our results, we prefer the (generally more conservative) estimates of the PCMH effect that include patient fixed effects. However, because these are within-patient estimates, they cannot control for factors like differential attrition which may potentially result from PCMH adoption and impact patient mix. Despite these qualifications, this paper offers expanded evidence the specific path to PCMH implementation matters, and calls for continued efforts to document and evaluate heterogeneity in primary care transformation.

4.6 Conclusion

One factor potentially driving the variation in estimates of the impact of the PCMH on patterns of utilization and spending is the substantial heterogeneity in how the model is implemented. The flexibility built into the NCQA guidelines may indeed allow for practices to tailor new health IT and other infrastructure to the needs of their individual patient populations. However, the evidence presented here and in previous work on heterogeneity in implementation suggests that the path practices take to PCMH recognition matters for patient utilization and spending outcomes. We find large reductions in spending with PCMH implementation, along with reduced ED visits. However, this reduction in ED utilization is concentrated

among a cluster of practices which emphasized the adoption of expanded electronic communications tools, potentially allowing patients to substitute communication with a primary care provider for an expensive trip to the ED. While further research is needed to document whether the reductions in utilization has any effect – positive or negative - on patient health, this work demonstrates the continued potential of the PCMH model and the need to carefully study the paths practices take to recognition.

References

Abadie, A., Drukker, D., Herr, J. L., & Imbens, G. W. (2004). Implementing matching estimators for average treatment effects in Stata. *The Stata Journal*, *4*(3), 290–311.

Adler, J. T., Yeh, H., Markmann, J. F., & Nguyen, L. L. (2016). Temporal analysis of market competition and density in renal transplantation volume and outcome. *Transplantation*, *100*(3), 670–677. https://doi.org/10.1097/TP.0000000000000851

Adler, P. S., & Clark, K. B. (1991). Behind the Learning Curve: A Sketch of the Learning Process. *Management Science*, *37*(3), 267–281. https://doi.org/10.1287/mnsc.37.3.267

Aggarwal, C., & Reddy, C. (2014). Hierarchical Clustering Algorithms. In *Data clustering : algorithms and applications*. Boca Raton, FL: CRC Press.

American Academy of Family Physicians, American Academy of Pediatrics, American College of Physicians, & American Osteopathic Association. (2007). *Joint Principles of the Patient-Centered Medical Home* . Retrieved from http://www.acponline.org/advocacy/where_we_stand/medical_home/approve_jp.pdf

Argote, L., & Epple, D. (1990). Learning Curves in Manufacturing. *Science*, *247*(4945), 920–924. https://doi.org/10.1126/science.247.4945.920

Arrow, K. J. (1962). The Economic Implications of Learning by Doing. *The Review of Economic Studies*, *29*(3), 155–173.

Autor, D. H. (2003). Outsourcing at Will: The Contribution of Unjust Dismissal Doctrine to the Growth of Employment Outsourcing. *Journal of Labor Economics*, *21*(1).

Axelrod, D. A., Guidinger, M. K., McCullough, K. P., Leichtman, A. B., Punch, J. D., & Merion, R. M. (2004). Association of center volume with outcome after liver and kidney transplantation. *American Journal of Transplantation*, *4*(6), 920–927. https://doi.org/10.1111/j.1600-6143.2004.00462.x

Bandi, C., Trichakis, N., & Vayanos, P. (2019). Robust Multiclass Queuing Theory for Wait Time Estimation in Resource Allocation Systems Robust Multiclass Queuing Theory for Wait Time

Estimation in Resource Allocation Systems. *Management Science*, *65*(1), 152–187.

Baroletti, S., & Dell'Orfano, H. (2010). Medication adherence in cardiovascular disease. *Circulation*, *121*(12), 1455–1458. https://doi.org/10.1161/CIRCULATIONAHA.109.904003

Beadles, C., Farley, J. F., Ellis, A. R., Lichstein, J. C., Morrissey, J. P., Dubard, C. A., & Domino, M. E. (2015). Do medical homes increase medication adherence for persons with multiple chronic conditions? *Medical Care*, *53*(2), 168–176. https://doi.org/10.1097/MLR.0000000000000292

Berenson, R. A., Hammons, T., Gans, D. N., Zuckerman, S., Merrell, K., Underwood, W. S., & Williams, A. F. (2008). A house is not a home: keeping patients at the center of practice redesign. *Health Affairs*, *27*(5), 1219–1230. https://doi.org/10.1377/hlthaff.27.5.1219

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). HOW MUCH SHOULD WE TRUST DIFFERENCES-IN-DIFFERENCES ESTIMATES? *The Quarterly Journal of Economics*, *119*(I), 249–275.

Bertsimas, D., Kung, J., Trichakis, N., Wojciechowski, D., & Vagefi, P. A. (2017). Accept or Decline ? An Analytics-Based Decision Tool for Kidney Offer Evaluation. *Transplantation*, *101*(12), 2898–2904. https://doi.org/10.1097/TP.0000000000001824

Billings, J., Parikh, N., & Mijanovich, T. (2000). *Emergency Department Use: The New York Story*.

Birkmeyer, J. D., Siewers, A. E., Finlayson, E. V. A., Stukel, T. A., Lucas, F. L., Batista, I., … Wennberg, D. E. (2002). Hospital Volume and Surgical Mortality in the United States. *New England Journal of Medicine*, *346*(15), 1128–1137. https://doi.org/10.1056/NEJMsa012337

Bodenheimer, T., Chen, E., & Bennett, H. D. (2009). Confronting the growing burden of chronic disease: Can the U.S. health care workforce do the job? *Health Affairs*, *28*(1), 64–74. https://doi.org/10.1377/hlthaff.28.1.64

Bodenheimer, T., & Pham, H. H. (2010). Primary care: Current problems and proposed solutions. *Health Affairs*, *29*(5), 799–805. https://doi.org/10.1377/hlthaff.2010.0026

Bodenheimer, T., Wagner, E. H., & Grumbach, K. (2002). Improving Primary Care for Patients With Chronic Illness. *Jama*, *288*(14), 1775–1779.

Borsky, A., Zhan, C., Miller, T., Ngo-Metzger, Q., Bierman, A. S., & Meyers, D. (2018). Few Americans receive all high-priority, appropriate clinical preventive services. *Health Affairs*, *37*(6), 925–928. https://doi.org/10.1377/hlthaff.2017.1248

Brown, R. S., Peikes, D., Peterson, G., Schore, J., & Razafindrakoto, C. M. (2012). Six Features Of Medicare Coordinated Care Demonstration Programs That Cut Hospital Admissions Of High-Risk Patients. *Health Affairs*, *31*(6), 1156–1166.

Buttorff, C., Ruder, T., & Bauman, M. (2017). *Multiple Chronic Conditions in the United States*. https://doi.org/10.7249/TL221

Cassidy, A. (2010). *Patient-Centered Medical Homes*. *Health Affairs*.

Cecka, J. M., Cohen, B., Rosendale, J., & Smith, M. (2006). Could more effective use of kidneys recovered from older deceased donors result in more kidney transplants for older patients? *Transplantation*, *81*(7), 966–970. https://doi.org/10.1097/01.tp.0000216284.81604.d4

Chandra, A., Gruber, J., & McKnight, R. (2011). The importance of the individual mandate--evidence from Massachusetts. *New England Journal of Medicine*, *364*(4), 293–295. https://doi.org/10.1056/NEJMp1013067

Christakis, N. A., & Lamont, E. B. (2000). Extent and determinants of error in doctors' prognoses in terminally ill patients: prospective cohort study. *British Medical Journal*, *320*, 469–473.

Christianson, J. B., Volmar, K. M., Alexander, J., & Scanlon, D. P. (2010). A Report Card on Provider Report Cards : Current Status of the Health Care Transparency Movement. *Journal of General Internal Medicine*, *25*(11), 1235–1241. https://doi.org/10.1007/s11606-010-1438-2

Cohen, J. B., Shults, J., Goldberg, D. S., Abt, P. L., Sawinski, D. L., & Reese, P. P. (2018). Kidney allograft offers: Predictors of turndown and the impact of late organ acceptance on allograft survival. *American Journal of Transplantation*, *18*(2), 391–401. https://doi.org/10.1111/ajt.14449

Conn, V. S., Ruppara, T. M., Enriqueza, M., & Cooper, P. (2016). Medication Adherence Interventions That Target Subjects with Adherence Problems: Systematic Review and Meta-analysis. *Research in Social and Administrative Pharmacy*, *12*(2), 218–246.

https://doi.org/10.1126/science.1249098.Sleep

Cotton, P. (2018). Patient-Centered Medical Home Evidence Increases With Time. Retrieved from https://www.healthaffairs.org/do/10.1377/hblog20180905.807827/full/

Currie, J., & MacLeod, W. B. (2017). Diagnosing Expertise: Human Capital, Decision Making, and Performance among Physicians. *Journal of Labor Economics*, *35*(1), 1–43. https://doi.org/10.1086/687848

David, G., Gunnarsson, C., Saynisch, P. A., Chawla, R., & Nigam, S. (2015). Do patient-centered medical homes reduce emergency department visits? *Health Services Research*, *50*(2), 418–439. https://doi.org/10.1111/1475-6773.12218

David, G., Saynisch, P. A., & Smith-McLallen, A. (2016). The Inner Workings of the Patient Centered Medical Home Model. *NBER Working Paper Series*, *22429*, 52.

David, G., Saynisch, P. A., & Smith-McLallen, A. (2018). The economics of patient-centered care. *Journal of Health Economics*, *59*, 60–77. https://doi.org/10.1016/j.jhealeco.2018.02.012

David, G., Saynisch, P., Luster, S., Smith-McLallen, A., & Chawla, R. (2018). The impact of patient-centered medical homes on medication adherence ? *Health Economics*, *27*, 1805–1820. https://doi.org/10.1002/hec.3804

David, G., Smith-McLallen, A., & Ukert, B. (2019). The Effect of Predictive Analytics-Driven Interventions on Healthcare Utilization. *Journal of Health Economics*. https://doi.org/10.1016/j.jhealeco.2019.02.002

Devries, A., Li, C. H., Sridhar, G., Hummel, J. R., Breidbart, S., & Barron, J. J. (2012). Impact of medical homes on quality, healthcare utilization, and costs. *Am J Manag Care*, *18*(9), 534–544. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/23009304

Deyo, R. A., Cherkin, D. C., & Ciol, M. A. (1992). Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *Journal of Clinical Epidemiology*, *45*(6), 613–619. https://doi.org/10.1016/0895-4356(92)90133-8

Domino, M.E., Humble, C., Lawrence Jr., W. W., & Wegner, S. (2009). Enhancing the medical homes

model for children with asthma. *Medical Care*, *47*(11), 1113–1120. https://doi.org/10.1097/MLR.0b013e3181adcc65

Domino, M.E., Wells, R., & Morrissey, J. P. (2015). Serving Persons With Severe Mental Illness in Primary Care–Based Medical Homes. *Psychiatric Services*, *66*(5), 477–483. https://doi.org/10.1176/appi.ps.201300546

Domino, Marisa Elena, Kilany, M., Wells, R., & Morrissey, J. P. (2017). Through the Looking Glass: Estimating Effects of Medical Homes for People with Severe Mental Illness. *Health Services Research*, *52*(5), 1858–1880. https://doi.org/10.1111/1475-6773.12585

Dranove, D., Kessler, D., McClellan, M., & Satterthwaite, M. (2003). Is More Information Better? The Effects of "Report Cards" on Health Care Providers. *Journal of Political Economy*, *111*(3), 555–588. https://doi.org/10.1086/374180

Emons, W. (1997). Credence goods and fraudulent experts. *The RAND Journal of Economics*, *28*(1), 107–119. https://doi.org/10.2307/2555942

Fox, C. R., & Tversky, A. (1998). A Belief-Based Account of Decision Under Uncertainty. *Management Science*, *44*(7), 879–895. https://doi.org/10.1287/mnsc.44.7.879

Franks, T. J., Burton, D. L., & Simpson, M. D. (2005). Patient medication knowledge and adherence to asthma pharmacotherapy: a pilot study in rural Australia. *Therapeutics and Clinical Risk Management*, *1*(1), 33–38. https://doi.org/10.2147/tcrm.1.1.33.53598

Friedberg, M. W., Lai, D. J., Hussey, P. S., & Schneider, E. C. (2009). A guide to the medical home as a practice-level intervention. *Am J Manag Care*, *15*(10 Suppl), S291-9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/20088633

Friedberg, M. W., Schneider, E. C., Rosenthal, M. B., Volpp, K. G., & Werner, R. M. (2014). Association between participation in a multipayer medical home intervention and changes in quality, utilization, and costs of care. *JAMA : The Journal of the American Medical Association*, *311*(8), 815–825. https://doi.org/10.1001/jama.2014.353

Gadkari, A. S., & McHorney, C. A. (2012). Unintentional non-adherence to chronic prescription

medications: How unintentional is it really? *BMC Health Services Research*, *12*(1). https://doi.org/10.1186/1472-6963-12-98

Gaynor, B. M., Seider, H., & Vogt, W. B. (2005). The Volume-Outcome Effect , Scale Economies , and Learning-by-Doing. *The American Economic Review*, *95*(2).

Green, L. A., Chang, H. C., Markovitz, A. R., & Paustian, M. L. (2018). The Reduction in ED and Hospital Admissions in Medical Home Practices Is Specific to Primary Care–Sensitive Chronic Conditions. *Health Services Research*, *53*(2), 1163–1179. https://doi.org/10.1111/1475-6773.12674

Halm, E. A., Lee, C., & Chassin, M. R. (2002). Is volume related to outcome in health care? A systematic review and methodologic critique of the literature. *Annals of Internal Medicine*, *137*(6), 511–520. https://doi.org/10.7326/0003-4819-137-6-200209170-00012

Hansen, C. B. (2007). Generalized least squares inference in panel and multilevel models with serial correlation and fixed effects. *Journal of Econometrics*, *140*, 670–694. https://doi.org/10.1016/j.jeconom.2006.07.011

Harbrecht, M. G., & Latts, L. M. (2012). Colorado's Patient-Centered Medical Home Pilot Met Numerous Obstacles, Yet Saw Results Such As Reduced Hospital Admissions. *Health Affairs*, *31*(9), 2010–2017. https://doi.org/10.1377/hlthaff.2012.0359

Hearld, L. R., & Alexander, J. A. (2012). Patient-centered care and emergency department utilization: a path analysis of the mediating effects of care coordination and delays in care. *Medical Care Research and Review*, *69*(5), 560–580. https://doi.org/10.1177/1077558712453618

Ho, V. (2002). Learning and the evolution of medical technologies: The diffusion of coronary angioplasty. *Journal of Health Economics*, *21*(5), 873–885. https://doi.org/10.1016/S0167-6296(02)00057-7

Hsiao, C.-J., & Hing, E. (2014). Use and characteristics of electronic health record systems among office-based physician practices: United States, 2001-2013. *NCHS Data Brief*, (143), 1–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/24439138

Huckman, R. S., & Pisano, G. P. (2006). The Firm Specificity of Individual Performance: Evidence from

Cardiac Surgery. *Management Science*, *52*(February 2015), 473–488. https://doi.org/10.1287/mnsc.1050.0464

Huesch, M. D. (2009). Learning by doing, scale effects, or neither? Cardiac surgeons after residency. *Health Services Research*, *44*(6), 1960–1982. https://doi.org/10.1111/j.1475-6773.2009.01018.x

Husain, S. A., Chiles, M. C., Lee, S., Pastan, S. O., Patzer, R. E., Tanriover, B., … Mohan, S. (2018). Characteristics and performance of unilateral kidney transplants from deceased donors. *Clinical Journal of the American Society of Nephrology*, *13*(1), 118–127. https://doi.org/10.2215/CJN.06550617

Imbens, G. W., Wooldridge, J. M., Blundell, R., Dias, M. C., Fraker, T., & Maynard, R. (2009). Recent Developments in the Econometrics of Program Evaluation. *Journal of Economic Literature*, *47*(1), 5–86.

Irwin, D. A., & Klenow, P. J. (1994). Learning-by-Doing Spillovers in the Semiconductor Industry. *Journal of Political Economy*, *102*(6), 1200–1227.

Jackson, G. L., Powers, B. J., Chatterjee, R., Bettger, J. P., Kemper, A. R., Hasselblad, V., … Williams, J. W. J. (2013). The Patient-Centered Medical Home: A Systematic Review. *Annals of Internal Medicine*, *158*(3), 169–178.

Jackson, G. L., & Williams, J. W. (2015). Does PCMH " Work "?— The Need to Use Implementation Science to Make Sense of Conflicting Results. *JAMA Internal Medicine*, *175*(8), 1369–1370. https://doi.org/10.1001/jamainternmed.2015

Joynt, K. E., Chan, D., Orav, E. J., & Jha, A. K. (2013). Insurance expansion in massachusetts did not reduce access among previously insured medicare patients. *Health Affairs*, *32*(3), 571–578. https://doi.org/10.1377/hlthaff.2012.1018

Ketchen Jr., D. J., & Shook, C. L. (1996). The Application of Cluster Analysis in Management Strategic Research : An Analysis and Critique. *Strategic Management Journal*, *17*(6), 441–458.

Khdour, M. R., Hawwa, A. F., Kidney, J. C., Smyth, B. M., & McElnay, J. C. (2012). Potential risk factors for medication non-adherence in patients with chronic obstructive pulmonary disease

(COPD). *European Journal of Clinical Pharmacology*, *68*(10), 1365–1373. https://doi.org/10.1007/s00228-012-1279-5

Kilo, C. M., & Wasson, J. H. (2010). Practice redesign and the patient-centered medical home: history, promises, and challenges. *Health Affairs*, *29*(5), 773–778. https://doi.org/10.1377/hlthaff.2010.0012

Koster, E., Philbert, D., Winters, N., & Bouvy, M. (2015). Adolescents' inhaled corticosteroid adherence: the importance of treatment perceptions and medication knowledge. *Journal of Asthma*, *52*(4).

Krass, I., Schieback, P., & Dhippayom, T. (2015). Adherence to diabetes medication: A systematic review. *Diabetic Medicine*, *32*(6), 725–737. https://doi.org/10.1111/dme.12651

Lai, J. C., Feng, S., & Roberts, J. P. (2012). An Examination of Liver Offers to Candidates on the Liver Transplant. *Gastroenterology*, *143*(5), 1261–1265. https://doi.org/10.1053/j.gastro.2012.07.105

Lapré, M. A., & Tsikriktsis, N. (2006). Organizational Learning Curves for Customer Dissatisfaction: Heterogeneity Across Airlines. *Management Science*, *52*(3), 352–366. https://doi.org/10.1287/mnsc.1050.0462

Lauffenburger, J. C., Shrank, W. H., Bitton, A., Franklin, J. M., Glynn, R. J., Krumme, A. A., … Choudhry, N. K. (2017). Association between patient-centered medical homes and adherence to chronic disease medications: A cohort study. *Annals of Internal Medicine*, *166*(2), 81–88. https://doi.org/10.7326/M15-2659

Lieberman, M. B. (1984). The Learning Curve and Pricing in the Chemical Processing Industries. *RAND Journal of Economics*, *15*(2), 213–228.

Lindquist, L. A., Go, L., Fleisher, J., Jain, N., Friesema, E., & Baker, D. W. (2012). Relationship of health literacy to intentional and unintentional non-adherence of hospital discharge medications. *Journal of General Internal Medicine*, *27*(2), 173–178. https://doi.org/10.1007/s11606-011-1886-3

Ludwig, J., Kling, J. R., & Mullainathan, S. (2011). Mechanism Experiments and Policy Evaluations. *Journal of Economic Perspectives*, *25*(3), 17–38. https://doi.org/10.1257/jep.25.3.17

Luft, H. S., Hunt, S. S., & Maerki, S. C. (1987). The volume-outcome relationship: practice-makes-perfect or selective-referral patterns? *Health Services Research*, *22*(2), 157–182.

Marrero, W. J., Naik, A. S., Friedewald, J. J., Xu, Y., Hutton, D. W., Lavieri, M. S., & Parikh, N. D. (2017). Predictors of Deceased Donor Kidney Discard in the United States. *Transplantation*, *101*(7), 1690–1697. https://doi.org/10.1097/TP.0000000000001238

Martsolf, G. R., Kandrack, R., Gabbay, R. A., & Friedberg, M. W. (2016). Cost of Transformation among Primary Care Practices Participating in a Medical Home Pilot. *Journal of General Internal Medicine*, *31*(7), 723–731. https://doi.org/10.1007/s11606-015-3553-6

McCarthy, D., Mueller, K., & Wrenn, J. (2009). *Geisinger Health System: Achieving the Potential of System Integration Through Innovation, Leadership, Measurement, and Incentives*. *Commonwealth Fund Case Study* (Vol. 9). https://doi.org/10.1016/S1360-8592(98)80013-2

Medicare.gov. (2016). How the Hospital Compare overall rating is calculated. Retrieved from https://www.medicare.gov/hospitalcompare/Data/Hospital-overall-ratings-calculation.html

Mishra, S. I., Gioia, D., Childress, S., Barnet, B., & Webster, R. L. (2011). Adherence to medication regimens among low-income patients with multiple comorbid chronic conditions. *Health and Social Work*, *36*(4), 249–258. https://doi.org/10.1093/hsw/36.4.249

Mojtabai, R., & Olfson, M. (2003). Medication Costs, Adherence, And Health Outcomes Among Medicare Beneficiaries. *Health Affairs*, *22*(4), 220–229. https://doi.org/10.1377/hlthaff.22.4.220

Monane, M., Bohn, R. L., Gurwitz, J. H., Glynn, R. J., & Avorn, J. (1994). Noncompliance with congestive heart failure therapy in the elderly. *Arch Intern Med*, *154*(4), 433–437.

National Committee for Quality Assurance. (2008). Standards and Guidelines for Physician Practice Connections® — Patient-Centered Medical Home (PPC-PCMH™). Washington, DC: National Committee for Quality Assurance.

National Committee for Quality Assurance. (2018). Patient-Centered Medical Home (PCMH). Retrieved from https://www.ncqa.org/programs/health-care-providers-practices/patient-centered-medical-home-pcmh/

NCQA. (2011). *Standards and Guidelines for NCQA's Patient - Centered Medical Home (PCMH) 2011*.

Nieuwlaat, R., Wilczynski, N., Navarro, T., Hobson, N., Jeffery, R., Keepanasseril, A., … Haynes, R.

(2008). Interventions for enhancing medication adherence (Review). *The Cochrane Database of Systematic Reviews*, (2), 2–4. https://doi.org/10.1002/14651858.CD000011.pub3.Copyright

Office of the National Coordinator for Health Information Technology. (2016). Office-based Physician Electronic Health Record Adoption. Retrieved from dashboard.healthit.gov/quickstats/pages/physician-ehr-adoption-trends.php

OPTN. (2018). National Data - OPTN. Retrieved November 19, 2018, from https://optn.transplant.hrsa.gov/data/view-data-reports/national-data/

Osterberg, L., & Blaschke, T. (2005). Adherence to medication. *The New England Journal of Medicine*, *353*(5), 487–497. https://doi.org/10.1056/NEJMra050100

Paquette, L., & Kida, T. (1988). The effect of decision strategy and task complexity on decision performance. *Organizational Behavior and Human Decision Processes*, *41*(1), 128–142. https://doi.org/10.1016/0749-5978(88)90051-9

Parekh, A. K., Goodman, R. A., Gordon, C., & Koh, H. K. (2011). Managing Multiple Chronic Conditions: A Strategic Framework for Improving Health Outcomes and Quality of Life. *Public Health Reports*, *126*(July-August), 460–471. https://doi.org/10.2307/41639389

Paustian, M. L., Alexander, J. A., El Reda, D. K., Wise, C. G., Green, L. A., & Fetters, M. D. (2013). Partial and Incremental PCMH Practice Transformation: Implications for Quality and Costs. *Health Services Research*. https://doi.org/10.1111/1475-6773.12085

Payne, J. W. (1976). Task complexity and contingent processing in decision making: An information search and protocol analysis. *Organizational Behavior and Human Performance*, *16*(2), 366–387. https://doi.org/10.1016/0030-5073(76)90022-2

Peikes, D., Dale, S., Lundquist, E., Genevro, J., Meyers, D., & Quality, A. for H. R. and. (2011). *Building the Evidence Base for the Medical Home: What Sample and Sample Size Do Studies Need? White Paper (Prepared by Mathematica Policy Research)*. Rockville, MD.

Peikes, D., Zutshi, A., Genevro, J. L., Parchman, M. L., & Meyers, D. S. (2012). Early evaluations of the medical home: building on a promising start. *Am J Manag Care*, *18*(2), 105–116. Retrieved from

http://www.ncbi.nlm.nih.gov/pubmed/22435838

Plumlee, M. A. (2003). The effect of information complexity on analysts' use of that information. *Accounting Review*, *78*(1), 275–296. https://doi.org/10.2308/accr.2003.78.1.275

Poznyak, D., Peikes, D. N., Wakar, B. A., Brown, R. S., & Reid, R. J. (2018). Development and Validation of the Modified Patient-Centered Medical Home Assessment for the Comprehensive Primary Care Initiative. *Health Services Research*, *53*(2), 944–973. https://doi.org/10.1111/1475-6773.12673

Punj, G., & Stewart, D. W. (1983). Cluster Analyis in Marketing Research: Review and Suggestions for Application. *Journal of Marketing Research*, *20*(2), 134–148.

Qaseem, A., Wilt, T. J., McLean, R. M., & Forciea, M. A. (2017). Noninvasive treatments for acute, subacute, and chronic low back pain: A clinical practice guideline from the American College of Physicians. *Annals of Internal Medicine*, *166*(7), 514–530. https://doi.org/10.7326/M16-2367

Reid, R. J., Coleman, K., Johnson, E. a, Fishman, P. a, Hsu, C., Soman, M. P., … Larson, E. B. (2010). The Group Health medical home at year two: cost savings, higher patient satisfaction, and less burnout for providers. *Health Affairs*, *29*(5), 835–843. https://doi.org/10.1377/hlthaff.2010.0158

Roebuck, M. C., Kaestner, R. J., & Dougherty, J. S. (2018). Impact of Medication Adherence on Health Services Utilization in Medicaid. *Medical Care*, *56*(3), 1. https://doi.org/10.1097/MLR.0000000000000870

Roebuck, M. C., Liberman, J., Gemmill-Toyama, M., & Brennan, T. (2011). Medication adherence leads to lower health care use and costs despite increased drug spending. *Health Affairs*, *30*(1), 91–99. https://doi.org/10.1377/hlthaff.2009.1087

Rosenthal, M. B., Alidina, S., Friedberg, M. W., Singer, S. J., Eastman, D., Li, Z., & Schneider, E. C. (2015). Impact of the Cincinnati Aligning Forces for Quality Multi-Payer Patient Centered Medical Home Pilot on Health Care Quality, Utilization, and Costs Meredith B. Rosenthal, Ph.D. *Medical Care Research and Review*, 1–22. https://doi.org/10.1177/1077558715618566

Rosenthal, M. B., Alidina, S., Friedberg, M. W., Singer, S. J., Eastman, D., Li, Z., & Schneider, E. C.

(2016). A Difference-in-Difference Analysis of Changes in Quality, Utilization and Cost Following the Colorado Multi-Payer Patient-Centered Medical Home Pilot. *Journal of General Internal Medicine*, *31*(3), 289–296. https://doi.org/10.1007/s11606-015-3521-1

Rosenthal, M. B., Friedberg, M. W., Singer, S. J., Eastman, D., Li, Z., & Schneider, E. C. (2013). Effect of a multipayer patient-centered medical home on health care utilization and quality: The rhode island chronic care sustainability initiative pilot program. *JAMA Internal Medicine*, *173*(20), 1907–1913. https://doi.org/10.1001/jamainternmed.2013.10063

Rosland, A. M., Wong, E., Maciejewski, M., Zulman, D., Piegari, R., Fihn, S., & Nelson, K. (2018). Patient-Centered Medical Home Implementation and Improved Chronic Disease Quality: A Longitudinal Observational Study. *Health Services Research*, *53*(4), 2503–2522. https://doi.org/10.1111/1475-6773.12805

Roth, A. E., Sönmez, T., & Ünver, M. U. (2004). Kidney exchange. *The Quarterly Journal of Economics*, *119*(2), 457–488. Retrieved from http://www.nber.org/papers/w10002

Roumie, C. L., Greevy, R., Wallston, K. A., Elasy, T. A., Kaltenbach, L., Kotter, K., … Speroff, T. (2011). Patient centered primary care is associated with patient hypertension medication adherence. *Journal of Behavioral Medicine*, *34*(4), 244–253. https://doi.org/10.1007/s10865-010-9304-6

Schold, J. D., Buccini, L. D., Srinivas, T. R., Srinivas, R. T., Poggio, E. D., Flechner, S. M., … Goldfarb, D. A. (2013). The association of center performance evaluations and kidney transplant volume in the united states. *American Journal of Transplantation*, *13*(1), 67–75. https://doi.org/10.1111/j.1600-6143.2012.04345.x

Shortell, S. M., Wu, F. M., Lewis, V. a., Colla, C. H., & Fisher, E. S. (2014). A taxonomy of accountable care organizations for policy and practice. *Health Services Research*, *49*(6), 1883–1899. https://doi.org/10.1111/1475-6773.12234

Sinaiko, B. A. D., Landrum, M. B., Meyers, D. J., Alidina, S., Maeng, D. D., Friedberg, M. W., … Rosenthal, M. B. (2017). Synthesis Of Research On Patient-Centered Medical Homes Brings Systematic Differences Into Relief. *Health Affairs*, *36*(3), 500–508.

Snowberg, E., & Wolfers, J. (2010). Explaining the Favorite – Long Shot Bias : Is it Risk-Love or Misperceptions ? *Journal of Political Economy*, *118*(4), 723–746.

Sokol, M. C., Mcguigan, K. A., Verbrugge, R. R., Robert, S., Care, S. M., Jun, N., … Epstein, R. S. (2005). Impact of Medication Adherence on Hospitalization Risk and Healthcare Cost. *Medical Care*, *43*(6), 521–530.

Sun, W., Wang, J., & Fang, Y. (2012). Regularized k-means clustering of high-dimensional data and its asymptotic consistency. *Electronic Journal of Statistics*, *6*(April 2011), 148–167. https://doi.org/10.1214/12-EJS668

Swietek, K. E., Domino, M. E., Beadles, C., Ellis, A. R., Farley, J. F., Grove, L. R., … Dubard, C. A. (2018). Do Medical Homes Improve Quality of Care for Persons with Multiple Chronic Conditions? *Health Services Research*, *53*(6), 4667–4681. https://doi.org/10.1111/1475-6773.13024

Tae, C. H., Jung, S. A., Moon, H. S., Seo, J. A., Song, H. K., Moon, C. M., … Jung, H. K. (2016). Importance of patients' knowledge of their prescribed medication in improving treatment adherence in inflammatory bowel disease. *Journal of Clinical Gastroenterology*, *50*(2), 157–162. https://doi.org/10.1097/MCG.0000000000000431

Thaler, R. H., & Ziemba, W. T. (2012). Anomalies: Parimutuel Betting Markets: Racetracks and Lotteries. *Journal of Economic Perspectives*, *2*(2), 161–174. https://doi.org/10.1257/jep.2.2.161

The New England Healthcare Institute. (2009). Thinking Outside the Pillbox: A System-wide Approach to Improving Patient Medication Adherence for Chronic Disease. *New England Healthcare Institute*, (August), 1–21. https://doi.org/10.1056/NEJMp1002305

Tirodkar, M. a, Morton, S., Whiting, T., Monahan, P., McBee, E., Saunders, R., & Scholle, S. H. (2014). There's more than one way to build a medical home. *Am J Manag Care*, *20*(12), e582-9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/25741875

UNOS. (2018). Transplant trends | UNOS. Retrieved November 19, 2018, from https://unos.org/data/transplant-trends/#waitlists_by_organ

USRDS. (2018a). *US Renal Data System Annual Data Report, Volume 2 Chapter 6*. Retrieved from

https://www.usrds.org/2018/view/v2_06.aspx

USRDS. (2018b). *US Renal Data System Annual Data Report, Volume 2 Chapter 9*. Retrieved from https://www.usrds.org/2017/view/v2_09.aspx

Varian, H. R. (2014). Big Data: New Tricks for Econometrics. *Journal of Economic Perspectives*, *28*(2), 3–28. https://doi.org/10.1257/jep.28.2.3

Verbeek, M., & Nijman, T. (1992). Testing for Selectivity Bias in Panel Data Models. *International Economic Review*, *33*(3), 681–703.

Verisk Health Inc. (2010). *Sightlines DxCG Risk Solutions: Concepts and Methodologies Guide for UNIX*. Waltham, MA.

Viswanathan, M., Golin, C. E., Jones, C. D., Ashok, M., Blalock, S., Wines, R. C. M., … Lohr, K. N. (2012). Closing the quality gap: revisiting the state of the science (vol. 4: medication adherence interventions: comparative effectiveness). *Evidence Report/Technology Assessment*, (208.4), 1–685. https://doi.org/10.1177/0145721712464400.Facilitating

Wagner, E. H., Austin, B. T., & Von Korff, M. (1996). Organizing care for patients with chronic illness. *Milbank Quarterly*, *74*(4), 511–544. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8941260

Ward, J. H. (1963). Hierarchical Grouping to Oprimize an Objective Function. *Journal of the American Statistical Association*, *58*(301), 236–244.

Weinick, R. M., Burns, R. M., & Mehrotra, A. (2010). Many emergency department visits could be managed at urgent care centers and retail clinics. *Health Affairs*, *29*(9), 1630–1636. https://doi.org/10.1377/hlthaff.2009.0748

Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology*, *39*(5), 806–820. https://doi.org/10.1037/0022-3514.39.5.806

Williams, R. C., Opelz, G., McGarvey, C. J., Weil, E. J., & Chakkera, H. A. (2016). The risk of transplant failure with hla mismatch in first adult kidney allografts from deceased donors. *Transplantation*, *100*(5), 1094–1102. https://doi.org/10.1097/TP.0000000000001115

Wooldridge, J. (2002). *Econometric Analysis of Cross Section and Panel Data.* Cambridge, MA: MIT

Press.

Wright, T. (1936). Factors Affecting the Cost of Airplanes. *Journal of the Aeronautical Sciences*, *3*. https://doi.org/10.1021/ie51396a026

Wu, G., & Gonzalez, R. (1999). Nonlinear Decision Weights in Choice Under Uncertainty. *Management Science*, *45*(1), 74–85. https://doi.org/10.1287/mnsc.45.1.74

Young, D. L., Goodie, A. S., Hall, D. B., & Wu, E. (2012). Decision making under time pressure, modeled in a prospect theory framework. *Organizational Behavior and Human Decision Processes*, *118*(2), 179–188. https://doi.org/10.1016/j.obhdp.2012.03.005

Appendix A - Appendix to Chapter 1

Table A1: Pr(Better Organ Within 6 Months of Turndown)

	LINEAR MODELS			LOGIT MODELS		
	(1)	(2)	(3)	(1)	(2)	(3)
ln(Lagged Registrants)	-0.0081*			-0.0998*		
	(0.0037)			(0.0416)		
ln(Lagged Cumul. Registrants)		-0.0015			-0.0276	
		(0.0034)			(0.0406)	
Registrant Q2			-0.0187*			-0.1525*
			(0.0071)			(0.0597)
Registrant Q3			-0.0294**			-0.2754***
			(0.0088)			(0.0764)
Registrant Q4			-0.0292**			-0.2723**
			(0.0103)			(0.0943)
N	758374	758374	758374	758374	758374	758374
R-sq	0.030	0.030	0.030	0.047	0.0467	0.0471
F	20.8276	20.7869	23.6843			

Standard errors in parentheses
+ p<0.1 * p<0.05 ** p<0.01 *** p<0.001

Appendix B - Appendix to Chapter 2

Figure B1: Level Description Venn Diagram

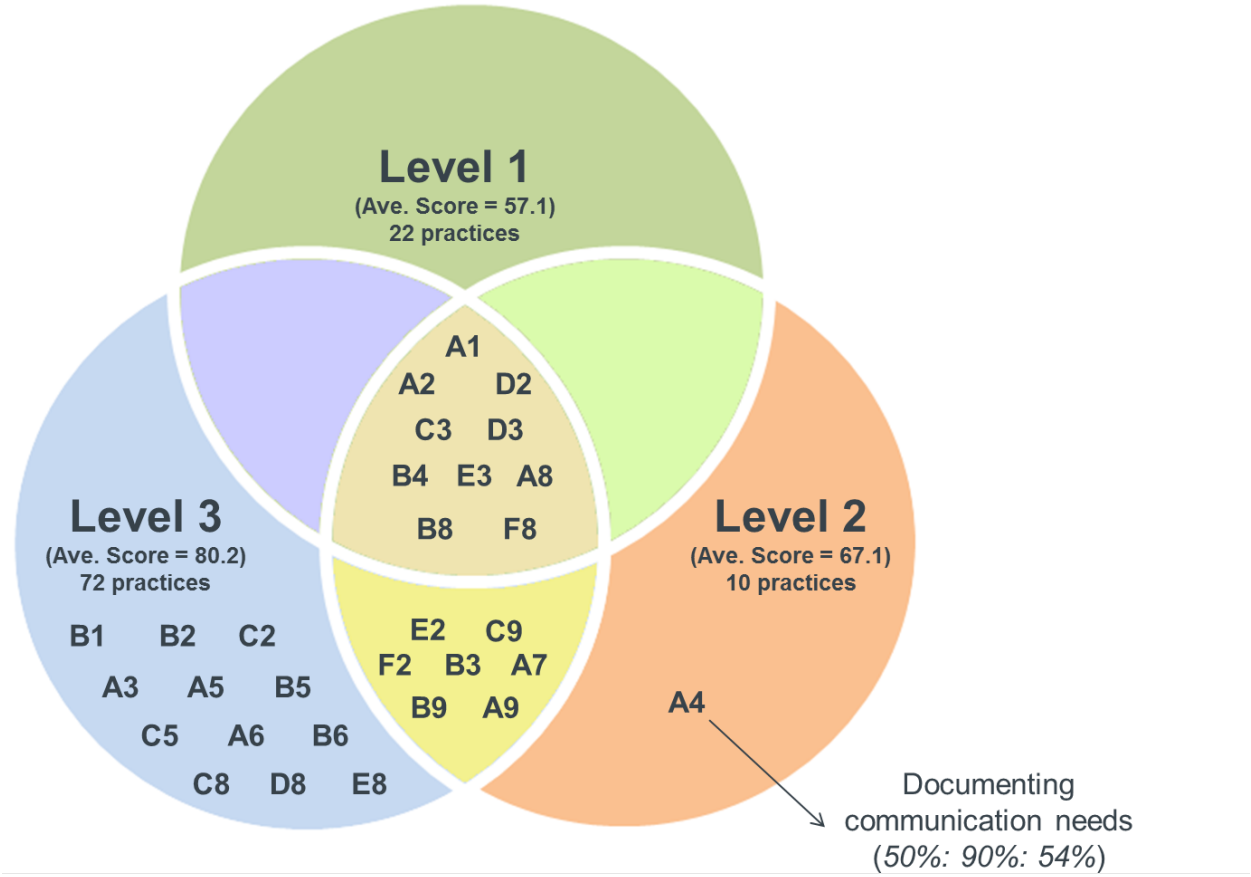


Table B1: 2008 NCQA PCMH Scoring Guidelines

Standards and Elements	Max Score
PPC1: Access and Communication	**9**
Access and communication processes**	4
Access and communication results**	5
PPC2: Patient Tracking and Registry Functions	**21**
Basic system for managing patient data	2
Electronic system for clinical data	3
Use of electronic clinical data	3
Organizing clinical data**	6
Identifying important conditions**	4
Use of system for population management	3
PPC3: Care Management	**20**
Guidelines for important conditions**	3
Preventive service clinician reminders	4
Practice organization	3
Care management for important conditions	5
Continuity of care	5
PPC4: Patient Self-Management Support	**6**
Documenting communication needs	2
Self-management support**	4

Standards and Elements	Max Score
PPC5: Electronic Prescribing	**8**
Electronic prescription writing	3
Prescribing decision support - safety	3
Prescribing decision support - efficiency	2
PPC6: Test Tracking	**13**
Test tracking and follow up**	7
Electronic system for managing tests	6
PPC7: Referral Tracking	**4**
Referral tracking**	4
PPC8: Performance Reporting and Improvement	**15**
Measures of performance**	3
Patient experience data	3
Reporting to physicians**	3
Setting goals and taking action	3
Reporting standardized measures	2
Electronic reporting to external entities	1
PPC9: Advanced Electronic Communications	**4**
Availability of interactive website	1
Electronic patient identification	2
Electronic care management support	1

**** Must Pass Element**

Table B2: ANOVA Results for Cluster Comparisons

NCQA Recognition Element	1	2	3	F-stat p-value
Access and communication processes**	89.39	81.50	100.00	3.236
	(4.22)	(5.04)	(0.00)	0.043
Access and communication results**	40.91	66.50	75.00	14.819
	(4.96)	(3.95)	(0.00)	0.000
Basic system for managing patient data	85.61	69.50	100.00	6.734
	(5.33)	(5.77)	(0.00)	0.002
Electronic system for clinical data	41.67	98.00	100.00	44.382
	(8.61)	(1.40)	(0.00)	0.000
Use of electronic clinical data	44.70	85.00	92.86	19.914
	(8.27)	(3.57)	(2.53)	0.000
Organizing clinical data**	93.18	99.00	100.00	2.009
	(4.25)	(1.00)	(0.00)	0.139
Identifying important conditions**	81.06	96.00	100.00	15.018
	(3.93)	(1.31)	(0.00)	0.000
Use of system for population management	33.33	70.50	50.00	15.639
	(5.18)	(5.04)	(0.00)	0.000
Guidelines for important conditions**	79.55	98.00	100.00	8.463
	(6.10)	(2.00)	(0.00)	0.000
Preventive service clinician reminders	62.88	93.00	100.00	21.534
	(6.63)	(2.03)	(0.00)	0.000
Practice organization	58.33	58.50	67.86	0.712
	(6.50)	(4.66)	(4.28)	0.493
Care management for important conditions	75.00	99.50	100.00	11.406
	(7.54)	(0.50)	(0.00)	0.000
Continuity of care	52.27	87.00	0.00	67.982
	(6.65)	(3.93)	(0.00)	0.000
Documenting communication needs	56.06	61.00	47.62	1.253
	(6.44)	(5.01)	(2.38)	0.290
Self-management support**	65.91	90.50	91.67	10.297
	(6.87)	(2.56)	(2.64)	0.000
Electronic prescription writing	46.21	80.00	92.86	12.675
	(8.08)	(4.84)	(4.92)	0.000
Prescribing decision support - safety	46.97	92.00	100.00	39.153
	(7.03)	(2.42)	(0.00)	0.000
Prescribing decision support - efficiency	44.70	94.00	100.00	42.546
	(7.60)	(1.96)	(0.00)	0.000
Test tracking and follow up**	31.82	84.00	100.00	59.601
	(5.99)	(3.33)	(0.00)	0.000
Electronic system for managing tests	44.70	96.00	100.00	53.417
	(7.20)	(1.31)	(0.00)	0.000
Referral tracking**	49.24	79.00	75.00	13.479
	(6.30)	(3.38)	(0.00)	0.000
Measures of performance**	78.79	90.00	100.00	3.362
	(6.89)	(4.04)	(0.00)	0.039
Patient experience data	78.79	75.00	100.00	3.406
	(6.89)	(6.10)	(0.00)	0.037
Reporting to physicians**	54.55	70.00	100.00	15.498
	(5.90)	(4.52)	(0.00)	0.000
Setting goals and taking action	31.82	74.00	50.00	16.597
	(6.08)	(5.39)	(0.00)	0.000
Reporting standardized measures	39.39	70.00	100.00	17.308
	(7.30)	(5.89)	(0.00)	0.000
Electronic reporting to external entities	33.33	35.50	100.00	23.216
	(6.94)	(6.55)	(0.00)	0.000
Availability of interactive website	18.18	53.00	75.00	23.069
	(5.36)	(5.33)	(0.00)	0.000
Electronic patient identification	0.00	14.00	0.00	4.269
	(0.00)	(4.96)	(0.00)	0.017
Electronic care management support	18.18	38.50	0.00	9.632
	(5.68)	(6.07)	(0.00)	0.000

Right-most column displays F-statistic from ANOVA comparison of three clusters

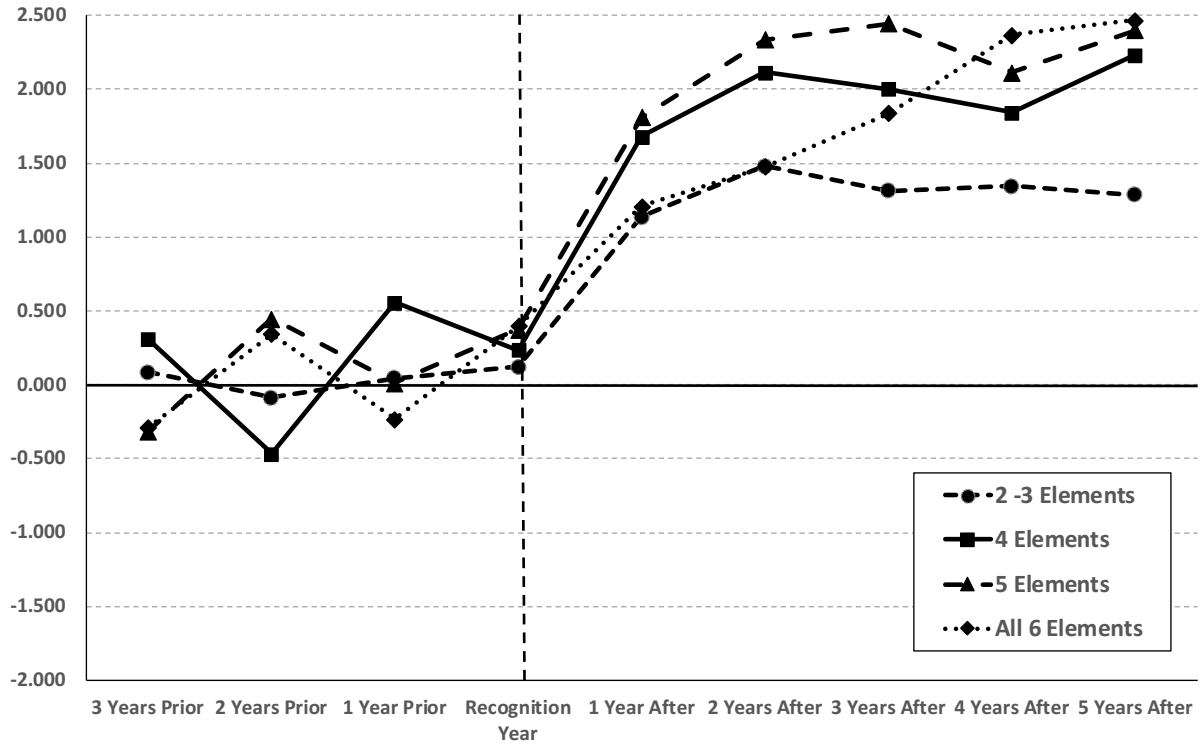
Table B3: Complete Results Table Using "Never Adopter" Practices as Controls

	Primary Care Visit		Specialist Visit		Hospitalization		Emergency Room Visit		Professional Expenditures	
	All	**Panel**	**All**	**Panel**	**All**	**Panel**	**All**	**Panel**	**All**	**Panel**
A - Baseline Model										
POST	**0.00678****	0.00144	-0.00074	-0.00565	-0.00019	0.00161	0.00471	0.00601	-5.404	-37.234
	[0.00311]	[0.00433]	[0.00293]	[0.00611]	[0.00241]	[0.00388]	[0.00371]	[0.00589]	[47.92]	[61.81]
Observations	3,399,446	1,128,377	3,399,446	1,128,377	3,399,446	1,128,377	3,399,446	1,128,377	2,868,564	976,046
R-squared	0.653	0.499	0.702	0.579	0.651	0.508	0.539	0.341	0.882	0.689
B - Level x POST interactions										
Level 1 x POST	0.00223	0.00209	-0.00096	-0.01157	-0.00256	0.00339	0.00110	0.00094	80.85	120.77
	[0.00657]	[0.00841]	[0.00695]	[0.0205]	[0.00447]	[0.00367]	[0.00622]	[0.00512]	[81.42]	[71.22]
Level 2 x POST	0.00161	0.00292	-0.00280	-0.00558	0.00275	0.00174	0.00139	0.00176	32.43	137.82
	[0.00598]	[0.00733]	[0.00708]	[0.00939]	[0.00412]	[0.00588]	[0.00619]	[0.00726]	[72.11]	[117.35]
Level 3 x POST	**0.00957*****	0.00153	-0.00042	-0.00087	-0.00324	-0.00118	0.00265	0.00215	-12.61	-54.21
	[0.00301]	[0.00578]	[0.00479]	[0.00591]	[0.00273]	[0.00415]	[0.00397]	[0.00512]	[32.44]	[58.92]
Observations	3,399,446	1,128,377	3,399,446	1,128,377	3,399,446	1,128,377	3,399,446	1,128,377	2,868,564	976,046
R-squared	0.655	0.508	0.702	0.579	0.651	0.508	0.539	0.340	0.880	0.689
C - Cluster x POST interactions										
Cluster 1 x POST	**-0.0573*****	**-0.0713*****	-0.0051	**-0.0141****	0.00311	0.00595	**0.0105*****	**0.01064****	-3.28	-17.19
	[0.00418]	**[0.00532]**	[0.00501]	**[0.00652]**	[0.00374]	[0.00495]	**[0.00393]**	**[0.00495]**	[59.33]	[72.33]
Cluster 2 x POST	**0.0256*****	**0.0250*****	0.00418	0.00134	-0.00012	0.00237	0.00354	0.00531	23.84	-16.38
	[0.00442]	**[0.00511]**	[0.00444]	[0.00581]	[0.00234]	[0.00412]	[0.00371]	[0.00497]	[48.77]	[62.51]
Cluster 3 x POST	**0.0266*****	**0.0243*****	**-0.0155*****	**-0.0176****	**-0.00597***	**-0.00768***	-0.00164	-0.00080	**-136.54****	**-134.11***
	[0.00398]	**[0.00541]**	**[0.00611]**	**[0.00713]**	**[0.00341]**	**[0.00426]**	[0.00511]	[0.00543]	**[59.69]**	**[71.74]**
Observations	3,399,446	1,128,377	3,399,446	1,128,377	3,399,446	1,128,377	3,399,446	1,128,377	2,868,564	976,046
R-squared	0.653	0.501	0.702	0.579	0.651	0.508	0.539	0.341	0.882	0.689

All regressions include member fixed effects. Standard errors in brackets: *** p<0.01, ** p<0.05, * p<0.1

Appendix C - Appendix to Chapter 3

Figure C1 – Comparison of Time Effects Pre- and Post-NCQA Recognition Year for Overall Adherence by Number of Elements Adopted in a Single Plot



Note: This figure overlays the four trends represented in Figure 3.2 to contrast trend-specific magnitudes across PCMH implementation with different number of adherence-related elements. The trend for 0 or 1 elements was dropped given the lack of significance and the corresponding magnitude of coefficient estimates.

Table C1 - American Hospital Formulary Service Codes

Group Description	AHFS Code	AHFS Description	Category
ARB	243208	ANGIOTENSIN II RECEPTOR ANTAGONISTS	ACE/ARB
ACE	243204	ANGIOTENSIN-CONVERTING ENZYME INHIBITORS	ACE/ARB
Antidepressants	281604	ANTIDEPRESSANTS	Antidepressants
Inhaled Corticosteroids	680400	ADRENALS	Asthma/COPD
Cromolyn	481032	MAST-CELL STABLILIZERS	Asthma/COPD
Leukotrienes	481024	LEUKOTRIENE MODIFIERS	Asthma/COPD
Monoclonal Antibodies	489200	RESPIRATORY TRACT AGENTS, MISCELLANEOUS	Asthma/COPD
Oral COPD	861600	RESPIRATORY SMOOTH MUSCLE RELAXANTS	Asthma/COPD
Beta Blocker	242400	BETA-ADRENERGIC BLOCKING AGENTS	Beta Blockers
Oral Hypoglycemics	682002	ALPHA-GLUCOSIDASE INHIBITORS	Diabetes
Oral Hypoglycemics	682004	BIGUANIDES	Diabetes
Oral Hypoglycemics	682005	DIPEPTIDYL PEPTIDASE-4(DPP-4) INHIBITORS	Diabetes
Oral Hypoglycemics	682006	INCRETIN MIMETICS	Diabetes
Oral Hypoglycemics	682016	MEGLITINIDES	Diabetes
Oral Hypoglycemics	682020	SULFONYLUREAS	Diabetes
Oral Hypoglycemics	682028	THIAZOLIDINEDIONES	Diabetes
Insulin	682008	INSULINS	Insulin
Lipid Lowering	240608	HMG-COA REDUCTASE INHIBITORS	Lipid
Lipid Lowering	240605	CHOLESTEROL ABSORPTION INHIBITORS	Lipid
Lipid Lowering	240606	FIBRIC ACID DERIVATIVES	Lipid
Lipid Lowering	240692	ANTILIPEMIC AGENTS, MISCELLANEOUS	Lipid

Table C2: Summary of PCMH Element Scoring/Coding

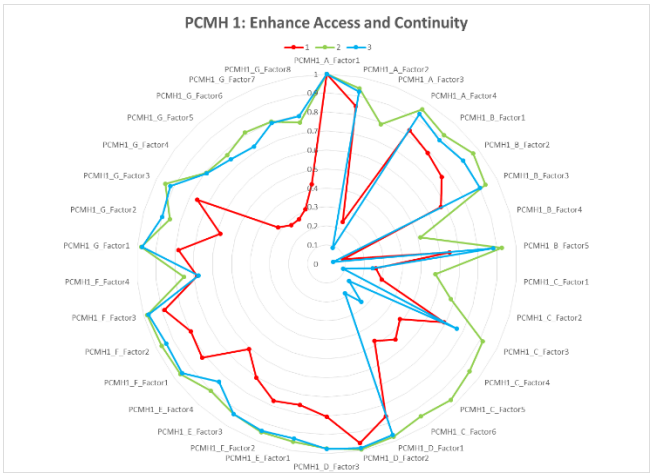
Element/ Factor	Description	NCQA Scoring	Coding
2D	*Organizing Clinical Data*	0% - Less than 10% of patient records include	Practices with a score greater than 50% were marked as "Y"; other practices were marked as "N."
		25% -10-24% of patient records seen in the past 3 months include at least 3 tools with information documented	
		50% - 25-49% of records of patients seen in the past 3 months include Factors 2 and 3 and at least 1 other with information documented	
		75% - 50-74% of records of patients seen in the past 3 months include Factors 2 and 3 and at least 1 other with information documented	
		100% - 75-100% of records of patients seen in the past 3 months include Factors 2 and 3 and at least 1 other with information documented.	
2F3	*Use of system for population management – specific medications*	The practice uses electronic information to generate lists of patients and take action to remind patients or clinicians proactively of services needed, as follows: Patients on a particular medication (Yes/No)	Practices designating "Yes" were marked as "Y"; practices designating "No" were marked as "N."
3D	*Care Management for Important Conditions*	0% - 10% or fewer patients seen in the past 3 months have at least 6 items documented	Practices with a score greater than 0% were marked as "Y"; other practices were marked as "N."
		25% - 11-24% of patients seen in the past 3 months have at least 6 items documented	
		50% - 25-49% of patients seen in the past 3 months have at least 6 items documented, including Factors 1, 3, 6, & 12 (for Tier II must also include Factor 2)	
		75% - 50-74% of patients seen in the past 3 months have at least 6 items documented, including Factors 1, 3, 6, & 12 (for Tier II must also include Factor 2)	
		100% - 75% or more of patients seen in the past 3 months have at least 6 items documented, including Factors 1, 3, 6, & 12 (for Tier II must also include Factor 2)	
5A	*Electronic prescription writing*	100% - 75-100% of new prescriptions for patients seen in the last 3 months written with Item 2 [Electronic prescription writer that is linked to patient-specific demographic and clinical information]	Practices where 75-100% of new prescriptions seen in the last three months were written with Item

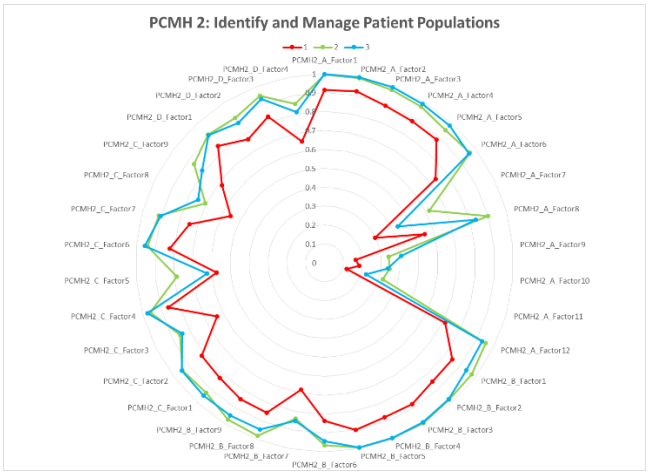
		75%- 75-100% of new prescriptions for patients seen in the last 3 months written with Item 1 [Electronic prescription writer-stand-alone system(general) with either print capability at the office or ability to send fax or electronic message to pharmacy]	2 (NCQA score = 100%), were designated "Y". All other practices were designated "N"
		25% -Practice has system capable of doing either Item 1 or Item 2, but practice does not use	
		0% -System does not have capability or less than 75% of prescriptions written with Item 1 or Item 2	
5B	*Prescribing decision support – safety*	100% - Practice uses 8 or more kinds of alerts and information	Practices using at least 2 kinds of alerts (NCQA scores of 50%, 75%, or 100%) were designated "Y". Practices with no capability, the capability of fewer than 6 kinds of alerts or using fewer than 2 kinds of alerts (NCQA score = 0%) were designated "N". The NCQA score of 25% was not present in our data set.
		75% - Practice uses 4 to 7 kinds of alerts and information	
		50% - Practice uses 2 to 3 kinds of alerts	
		25% - System has capability of providing 6 or more kinds of alerts, but practice does not use them.	
		0% - No system capability, system has capability for fewer than 6 kinds of alerts or practice uses fewer than 2 kinds of alerts and information	
5C	*Prescribing decision support – efficiency*	100% - Practice uses 2 tools [Electronic prescriptions writer with general automatic alerts for different choices including generics AND Electronic prescription writer connected to payer-specific formulary that automatically alerts clinician to alternative drugs, including generics.]	Practices using at least one tool (NCQA scores of 75% or 100%) were designated "Y", all other practices were designated "N".
		75% - Practice uses 1 tool	
		25% - System has capability to support both options; practice does not use it	
		0% System does not have capability or practice does not use either tool	

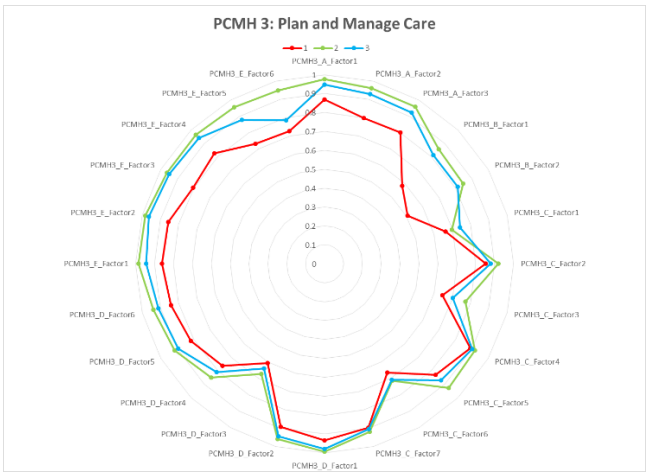
The NCQA often scores a particular element as 0%, 25%, 50%, 75% or 100%. What those percentages actually mean differs by the element. Thus, in the NCQA scoring column, the value to the left of the dash is the NCQA score, with the description of what that score means on the right.

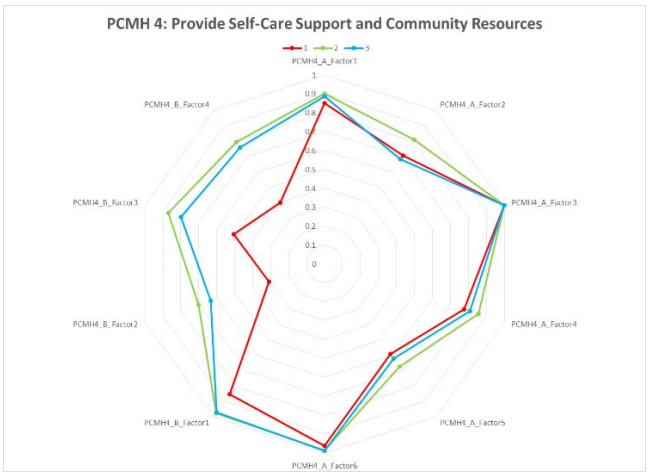
Appendix D - Appendix to Chapter 4

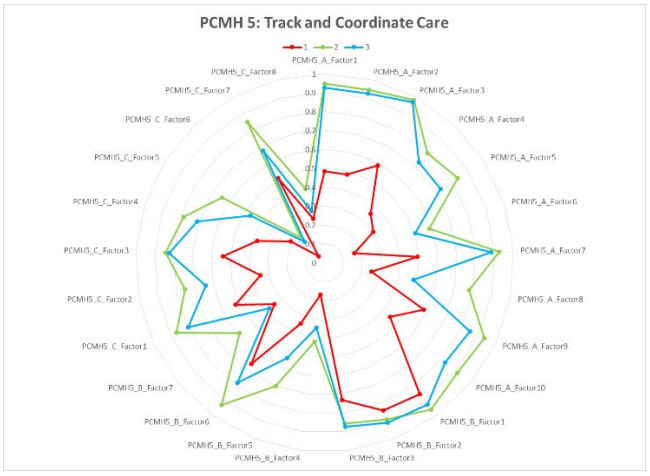
Figure D1: Factor-Level Radar Plots











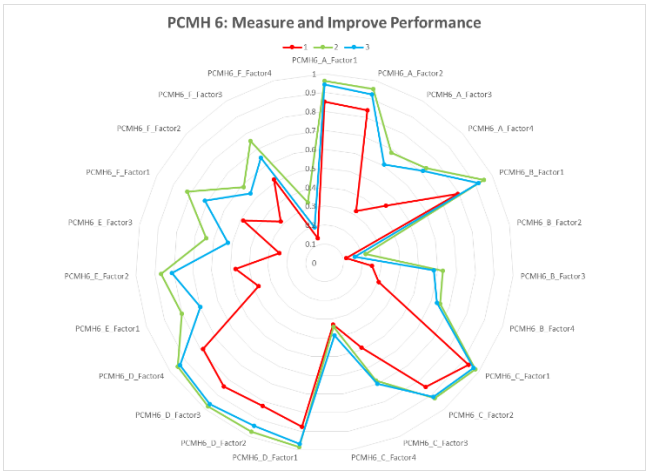


Table D1: One-way ANOVA Comparisons of Cluster Performance, Element Level

Element	Cluster 1	Cluster 2	Cluster 3	p-value, 1 vs 2 vs 3	p-value, 2 vs 3
1A - Access During Office Hours*†	82.27	93.71	90.16	0.00000	0.00000
1B - After-Hours Access†	59.76	83.63	81.17	0.00000	0.00001
1C - Electronic Access†	62.13	95.51	49.50	0.00000	0.00000
1D - Continuity	83.28	97.26	96.67	0.00000	0.05048
1E - Medical Home Responsibilities†	70.93	93.41	91.11	0.00000	0.00000
1F - Culturally and Linguistically Appropriate Services†	85.36	97.55	96.01	0.00000	0.00000
1G - The Practice Team	42.39	82.43	83.49	0.00000	0.13489
2A - Patient Information	74.22	93.17	92.66	0.00000	0.14310
2B - Clinical Data†	82.79	97.43	95.35	0.00000	0.00000
2C - Comprehensive Health Assessment	77.61	93.01	93.08	0.00000	0.86854
2D - Use Data for Population Management*†	77.12	90.48	88.05	0.00000	0.00000
3A - Implement Evidence-Based Guidelines†	79.16	95.54	91.60	0.00000	0.00000
3B - Identify High-Risk Patients†	51.52	84.17	80.13	0.00000	0.00000
3C - Care Management*†	81.67	88.52	86.32	0.00000	0.00000
3D - Medication Management†	87.09	94.89	92.73	0.00000	0.00000
3E - Use Electronic Prescribing†	82.85	98.11	94.01	0.00000	0.00000
4A - Support Self-Care Process*†	88.09	93.98	90.81	0.00007	0.00000
4B - Provide Referrals to Community Resources†	51.75	83.58	79.14	0.00000	0.00000
5A - Test Tracking and Follow-Up†	36.78	91.92	85.33	0.00000	0.00000
5B - Referral Tracking and Follow-Up†	72.22	92.77	86.10	0.00000	0.00000
5C - Coordinate With Facilities/Care Transitions*†	60.72	92.47	84.60	0.00000	0.00000
6A - Measure Performance†	70.06	87.25	84.57	0.00000	0.00000
6B - Measure Patient/Family Experience†	37.21	61.60	57.70	0.00000	0.00000
6C - Implement Continuous Quality Improvement*	74.03	86.09	86.80	0.00000	0.19493
6D - Demonstrate Continuous Quality Improvement†	83.33	97.20	95.18	0.00000	0.00000
6E - Report Performance†	40.67	80.54	72.19	0.00000	0.00000
6F - Report Data Externally†	65.70	92.69	85.83	0.00000	0.00000

Table D2: Overall PCMH Adoption Effect on Log-transformed Expenditures, by Category

	Patient FE	Practice FE
Total Costs	-0.0833***	-0.111***
	(0.00502)	(0.0177)
Inpatient	0.00261	0.00162
	(0.00650)	(0.00706)
Emergency Department	-0.0177***	-0.0233***
	(0.00663)	(0.00790)
General physician	-0.106***	-0.159***
	(0.00579)	(0.0234)
Specialty	-0.0562***	-0.0770***
	(0.00613)	(0.0147)
Imaging service	-0.0464***	-0.0542***
	(0.00733)	(0.0143)
Lab service	-0.0837***	-0.105***
	(0.00561)	(0.0205)

Reported coefficients are for overall PCMH indicator. Results exclude observations with partial patient-year or which occurred during the PCMH recognition year. Regression model also adjusted for year (2006-2016), age, age-squared, gender, comorbidity index, COPD, CHF, malignancy, and diabetes.

*** p<0.01, ** p<0.05, * p<0.1

Table D3: PCMH Adoption Effect Estimates, Intensive Margin

	Patient FE	Practice FE
Inpatient visit	0.0148	0.0127
	(0.0155)	(0.00905)
ED visit	0.0274**	0.0123*
	(0.0109)	(0.00660)
General physician visit	-0.0337***	-0.0395***
	(0.00607)	(0.0114)
Specialty visit	-0.0543***	-0.0439***
	(0.0119)	(0.0161)
Imaging service	-0.0168	-0.0228
	(0.0142)	(0.0198)
Lab service	-0.00801	-0.00770
	(0.0140)	(0.0331)

Reported coefficients are for overall PCMH indicator. Results exclude observations with partial patient-year or which occurred during the PCMH recognition year. Regression model also adjusted for year (2006-2016), age, age-squared, gender, comorbidity index, COPD, CHF, malignancy, and diabetes.
*** p<0.01, ** p<0.05, * p<0.1

Table D4: PCMH Adoption Effect by Cluster on Log-Transformed Expenditures, by Category

Total Costs	Patient FE	Practice FE	Outpatient	Patient FE	Practice FE	Imaging	Patient FE	Practice FE
Cluster 1 x Post	-0.0868***	-0.131***	Cluster 1 x Post	-0.134***	-0.0909***	Cluster 1 x Post	-0.0680***	-0.0852***
	(0.00870)	(0.0353)		(0.0184)	(0.00830)		(0.0124)	(0.0274)
Cluster 2 x Post	-0.0829***	-0.106***	Cluster 2 x Post	-0.109***	-0.0841***	Cluster 2 x Post	-0.0426***	-0.0486***
	(0.00513)	(0.0149)		(0.00835)	(0.00488)		(0.00748)	(0.0135)
Cluster 3 x Post	-0.0836***	-0.129***	Cluster 3 x Post	-0.133***	-0.0885***	Cluster 3 x Post	-0.0582***	-0.0721***
	(0.00765)	(0.0353)		(0.0167)	(0.00730)		(0.0108)	(0.0239)
Inpatient	**Patient FE**	**Practice FE**	**General Physician**	**Patient FE**	**Practice FE**	**Labs**	**Patient FE**	**Practice FE**
Cluster 1 x Post	0.0154	0.00958	Cluster 1 x Post	-0.148***	-0.199***	Cluster 1 x Post	-0.0617***	-0.0762*
	(0.0113)	(0.0128)		(0.0106)	(0.0492)		(0.0101)	(0.0410)
Cluster 2 x Post	0.00171	0.000939	Cluster 2 x Post	-0.105***	-0.156***	Cluster 2 x Post	-0.0827***	-0.103***
	(0.00661)	(0.00710)		(0.00594)	(0.0231)		(0.00573)	(0.0203)
Cluster 3 x Post	0.000476	0.00105	Cluster 3 x Post	-0.0809***	-0.149***	Cluster 3 x Post	-0.105***	-0.136***
	(0.00949)	(0.0106)		(0.00906)	(0.0442)		(0.00865)	(0.0331)
Emergency Department	**Patient FE**	**Practice FE**	**Specialist**	**Patient FE**	**Practice FE**			
Cluster 1 x Post	-0.00735	-0.00423	Cluster 1 x Post	-0.0713***	-0.122***			
	(0.0111)	(0.0148)		(0.0108)	(0.0293)			
Cluster 2 x Post	-0.0218***	-0.0284***	Cluster 2 x Post	-0.0580***	-0.0719***			
	(0.00674)	(0.00826)		(0.00627)	(0.0141)			
Cluster 3 x Post	0.00408	-0.00156	Cluster 3 x Post	-0.0342***	-0.0816***			
	(0.00966)	(0.0147)		(0.00937)	(0.0292)			

Reported coefficients are for Cluster x Post indicators. Results exclude observations with partial patient-year or which occurred during the PCMH recognition year.
Regression model also adjusted for year (2006-2016), age, age-squared, gender, comorbidity index, COPD, CHF, malignancy, and diabetes.
*** p<0.01, ** p<0.05, * p<0.1