NBER WORKING PAPER SERIES

THE INNER WORKINGS OF THE PATIENT CENTERED MEDICAL HOME MODEL

Guy David
Philip A. Saynisch
Aaron Smith-McLallen

Working Paper 22429
http://www.nber.org/papers/w22429

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2016

We thank the participants of seminars at University of Illinois at Urbana Champaign, Harvard, University of Minnesota, University of Illinois at Chicago, and University of Pennsylvania. As well as participants of the 2016 American Health Economics Conference, the 6th Biennial Conference of the American Society of Health Economists, and the IDC School of Economics conference. Special thanks to Ravi Chawla, Somesh Nigam, Cindy Wang, Sarah Scholle, and Candace Gunnarsson for data support and helpful suggestions. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed a financial relationship of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w22429.ack

The Inner Workings of the Patient Centered Medical Home Model
Guy David, Philip A. Saynisch, and Aaron Smith-McLallen
NBER Working Paper No. 22429
July 2016
JEL No. I1,I10,I13,I18,L21,L23

ABSTRACT

The patient-centered medical home (PCMH) is a model for restructuring primary care with a focus on improved access to care and clinical excellence. However, to date, the evidence on its effect on healthcare utilization and expenditures has been quite mixed. One possible reason for this may lie in the flexibility with which a practice can meet the thresholds for PCMH recognition, adopting practice reforms in vastly different domains. Hence, practices with the same recognition level may in fact demonstrate divergent capabilities, and subsequently have different leverage over the achievement of the PCMH goals. We study this idea by using previously unavailable data that spans all PCMH recognition domains. The richness of our data allows us to group practices into clusters based on their choice of attributes during the recognition process, and then examine the performance of individual clusters in reducing healthcare utilization and expenditures. We find that treating the PCMH model as an undifferentiated intervention obscures meaningful variation in implementation across practices. In particular, clusters emphasizing practice improvements like use of decision support, enhanced access to care and population management tools have demonstrated some success in impacting utilization and expenditures patterns after PCMH recognition.

Guy David
The Wharton School
University of Pennsylvania
202 Colonial Penn Center
3641 Locust Walk
Philadelphia, PA 19104-6218
and NBER
gdavid2@wharton.upenn.edu

Philip A. Saynisch
Harvard Business School
psaynisch@hbs.edu

Aaron Smith-McLallen
Independence Blue Cross
1900 Market Street
Philadelphia, PA 19103-1480
Aaron.Smith-McLallen@ibx.com

I. Introduction

The increasing prevalence of chronic illnesses and mounting associated costs are major concerns for the US healthcare system. Along with other reforms, considerable attention has been devoted to efforts to restructure primary care with the goal of improving the coordination of care. One model for reform which has generated considerable attention is the Patient-Centered Medical Home (PCMH), 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 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 is extremely detailed, including scoring based on 127 individual "factors" of implementation.[1] 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 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. Using this approach, we find substantial heterogeneity in implementation across clusters, with different areas of focus (generally summarized as "the basic model", "enhanced population health management" and "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

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

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 III introduces a theoretical model of profit or value-driven procurement under asymmetric information to describe the contractual interactions between insurers and primary care practices. The model highlights the role of quality assurance credentialing (such as NCQA's PCMH recognition process), which certifies that practices have adopted a range of quality-improving features. Some of these features may have cost-reducing effects valued by the insurer, thereby providing a noisy signal of efficiency to insurers. This certification process may play a role in mitigating adverse selection and in ensuring that certain productive tasks are performed by the practice.

Having provided a theoretical framework for understanding why insurers may value the signal of quality provided by NCQA accreditation, we then assess how the signal is being produced. Whereas previous analyses have been limited to coarse summaries of PCMH recognition (either a binary indicator of PCMH adoption or the level of recognition), this study makes use of the granular practice data used by NCQA in the recognition process. Section IV discusses the sources of patient and primary care practice data used in this study. Section V describes our empirical approach, covering both the clustering technique and regression analysis used to assess patient-level data. In Section VI, 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 VII includes discussion of these results, and Section VIII concludes.

II. The Patient Centered Medical Home

The status quo approach to organizing primary care is poorly suited to address the needs of patients. An extensive body of literature points to a number of problems - 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.[2]

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 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

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

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 Table 1.

[TABLE 1]

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. While 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., 2014; Rosenthal et al., 2015, 2016), other studies, 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.

III. Theoretical Model

The purpose of this section is to describe the potential value of NCQA recognition to insurers in a principal-agent procurement framework. In this framework, asymmetric information allows for informational rents and distortion of optimal levels of an unobserved efficiency parameter. There is an extensive literature on the value and desirability of auditing (Khalil & Lawarree, 1995), supervision (Engel, 2006) and monitoring (Lewis & Sappington, 1991; Liu, 1982; Zhao, 2008) in reducing such distortions. This literature emphasizes threats of collusion between the supervisor and either party - principal or agent (Strausz, 1997; Vafaï, 2005). For example, Strausz (1997) shows the delegation of monitoring to a third-party supervisor is profitable even in the face of collusion. While most of this literature analyzes these issues in the case of within-firm hierarchical labor relations between owners and managers (or managers and their staff), our context lends itself to a procurement framework, as it pertains to cases were the insurer is not financially integrated with the physician practice with which it contracts.[3] Since a health care provider must consent to any type of external monitoring, we focus on information extraction by the insurer that relies on third-party credentialing organizations (such as NCQA). We disregard the possibility of collusion between NCQA and health care providers, but instead point to the fact that there may be a misalignment between the parameters collected by the NCQA and the ones the insurer cares about. In particular, the information sought by a profit maximizing insurer is likely to differ from the information generated through a process designed to achieve social value. Moreover, based on the underlying NCQA process, recognition can be achieved through different combinations of actions (or clusters of actions). And while these different clusters can plausibly lead to the same desirable social outcome, they may differentially affect the insurer's profitability.

Assume a risk-neutral insurer and a risk-neutral primary care practice which respectively insure and serve a group of individuals. Extending the simple principal-agent framework, we assume that the practice can

[3] Notable exceptions include integrated delivery systems that have an insurance component (typically an HMO) under a single parent company. Examples include Kaiser Permanente in California, Geisinger Health System in Pennsylvania, or Intermountain Healthcare in Utah.

take on a set of indivisible actions not fully observed by the insurer. A specific combination of actions would yield a value, π, to a pure profit maximizing insurer. We assume that π is large enough to justify the creation of a contractual environment that would induce the practice to take such actions. In addition, we assume that the sole motivation for the practice in taking an action is to maximize the difference between the monetary transfer from the insurer, *t*, and its cost of taking such action. This excludes the possibility that these actions benefit the practice in other ways, in particular, the practice may derive utility from enhancing its quality and reputation either due to altruistic motives or through its ability to translate quality reputation into surplus (e.g. by charging higher prices, increasing its patient panel, etc.).

For simplicity, we will consider three specific actions taken by the practice. The first is an action that benefits both patients and the insurer, implemented at effort level e_{ip}. For example, the practice can allocate resources to improve population health management which results in a reduction in avoidable downstream health care utilization. The second is an action that benefits only patients, with effort expressed as e_p. For example, the practice may refer patients for treatment that is both more effective and more expensive. The third is an action that benefits only the insurer, with effort level, e_i. For example, referring a patient to an equally competent, yet lower cost specialist. The total level of effort is the sum of the three action components (i.e. $\vec{e} = e_p + e_i + e_{ip}$).

Since these actions may require different diagnostic, technological, and informational capabilities the practice may come to realize that it is well suited for some actions and less so for others. This concept is modeled by variation in the efficiency vector $\vec{\theta}$, which is realized after a contract is offered by the insurer to the practice. When the practice is efficient in performing an action, high effort has relatively low disutility associated with it. The efficiency vector is represented by a set of efficiency parameters that correspond to the actions in the effort vector (i.e. $\vec{\theta} = \theta_p + \theta_i + \theta_{ip}$).

The cost of effort (or a measure of disutility of effort in monetary terms) is assumed to be additively separable in the three effort components and proportional to the parameters in the efficiency vector, such that: $\gamma(\vec{\theta},\vec{e}) = \frac{1}{2}\theta_p^{-1}\cdot e_p^2 + \frac{1}{2}\theta_i^{-1}\cdot e_i^2 + \frac{1}{2}\theta_{ip}^{-1}\cdot e_{ip}^2$. Letting subscripts denote partial derivatives, for any component of *e*, we verify that $\gamma(\vec{\theta},0)=0$, $\gamma_e(\vec{\theta},\vec{e})\geq 0$, $\gamma_{ee}(\vec{\theta},\vec{e})>0$, $\gamma_\theta(\vec{\theta},\vec{e})<0$, and $\gamma_{e\theta}(\vec{\theta},\vec{e})\leq 0$. This cost function implies that the marginal disutility from effort at any effort level, $\gamma_e(\vec{\theta},\vec{e})$ is lower when $\vec{\theta}$ is higher. For example, reducing medication errors is easier when the practice uses an electronic system to order prescriptions, to check for safety and to promote efficiency when prescribing. Similarly, identifying patient language preferences to overcome patient-specific barriers to communication is easier when the practice staff is fluent in that language. Finally, we focus on the illustratively and computationally simpler case in which each component θ_j, where $j\in\{p,i,ip\}$, can take only one of two values, $\theta_{j,H}$ or $\theta_{j,L}$, with $\theta_{j,H}>\theta_{j,L}>0$. Since $\vec{\theta}$ is realized after the contract is introduced, we assume that the insurer knows the true distribution of each individual component of $\vec{\theta}$, such that $\text{Prob}(\theta_{j,H}) = q_j \in (0,1)$.

The insurer has access to a signal, *s*, on $\vec{\theta}$, and can refine its contracts to the practice based on this signal. The information available to the insurer comes from a third-party credentialing organization through the implementation of a quality assurance initiative (e.g. NCQA PCMH recognition). The signal is a discrete credentialing level – a new information that allows for the screening of practices to more effectively combat adverse selection.[4] Furthermore, note that the third-party credentialing organization only observes the realization of capabilities, θ, but not how they are used, *e*. It is interested solely in quality of care, and therefore collects information exclusively on θ_p and θ_{ip}.

[4] Third-party screening, while producing a signal for the insurer should be distinguished from the concept of *Signaling*, which implies that the informed practice moves first by choosing an action to signal its type.

Therefore, after the contract is set, the insurer observes the credentialing level s_H when $\bar{s} \le \theta_p + \theta_{ip}$ or s_L when $\bar{s} > \theta_p + \theta_{ip}$ that the practice obtained. Since the profit maximizing insurer does not care about θ_p and the signal contains no information about θ_i, its informational value depends exclusively on how θ_p and θ_{ip} map into *s*. To make credentialing instructive in this setting let $\bar{s} < \theta_{p,H} + \theta_{ip,H}$ and $\bar{s} > \theta_{p,L} + \theta_{ip,L}$, that is, s_H and s_L correctly correspond to the realizations (H,H) and (L,L). Therefore, the key question is how $\theta_{p,L} + \theta_{ip,H}$ and $\theta_{p,H} + \theta_{ip,L}$ map onto *s*. If $\theta_{p,L} + \theta_{ip,H}$ belongs to s_H and $\theta_{p,H} + \theta_{ip,L}$ to s_L, the signal accurately reveals the state of θ_{ip}.

Under this scenario, the insurer observes θ_{ip} but not θ_i, and hence can specify two conditional contracts $((t_{HH}, e_{i,H}), (t_{HL}, e_{i,H}) \mid s_H)$ and $((t_{LH}, e_{i,L}), (t_{LL}, e_{i,L}) \mid s_L)$ that take into account the random nature of θ_i and condition on the realization of θ_{ip}. The two resulting individual rationality (IR) constraints will be a hybrid between the case where all values of θ are observed and the case where none of the values of θ are observed. The IR constraints are presented in III.2 and III.3. The incentive compatibility (IC) constraints (III.4 and III.5) ensure that a contract designed for one type of practice would be preferred by that practice to the one designed for the other practice type.

The expected profit maximization problem for the insurer is given by:

$$\underset{t_{HH},t_{HL},t_{LH},t_{LL}}{Max} \quad \begin{aligned} E\pi = {} & q_i \cdot q_{ip} \cdot (e_{i,H} + e_{ip,H} - t_{HH}) + q_i \cdot (1-q_{ip}) \cdot (e_{i,H} + e_{ip,L} - t_{HL}) + \\ & + (1-q_i) \cdot q_{ip} \cdot (e_{i,L} + e_{ip,H} - t_{LH}) + (1-q_i) \cdot (1-q_{ip}) \cdot (e_{i,L} + e_{ip,L} - t_{LL}) \end{aligned} \qquad \text{(III.1)}$$

$$\textbf{s.t.} \qquad q_{ip} \cdot \left(t_{LH} - \frac{e_{i,L}^2}{2\theta_{i,L}} - \frac{e_{ip,H}^2}{2\theta_{ip,H}}\right) + (1-q_{ip}) \cdot \left(t_{LL} - \frac{e_{i,L}^2}{2\theta_{i,L}} - \frac{e_{ip,L}^2}{2\theta_{ip,L}}\right) \ge 0 \qquad \text{(III.2)}$$

$$q_{ip} \cdot (t_{HH} - \frac{e_{i,H}^2}{2\theta_{i,H}} - \frac{e_{ip,H}^2}{2\theta_{ip,H}}) + (1-q_{ip}) \cdot (t_{HL} - \frac{e_{i,H}^2}{2\theta_{i,H}} - \frac{e_{ip,L}^2}{2\theta_{ip,L}}) \geq 0 \qquad \text{(III.3)}$$

$$\begin{aligned} &q_{ip} \cdot (t_{HH} - \frac{e_{i,H}^2}{2\theta_{i,H}} - \frac{e_{ip,H}^2}{2\theta_{ip,H}}) + (1-q_{ip}) \cdot (t_{HL} - \frac{e_{i,H}^2}{2\theta_{i,H}} - \frac{e_{ip,L}^2}{2\theta_{ip,L}}) \geq \\ &q_{ip} \cdot (t_{LH} - \frac{e_{i,L}^2}{2\theta_{i,H}} - \frac{e_{ip,H}^2}{2\theta_{ip,H}}) + (1-q_{ip}) \cdot (t_{LL} - \frac{e_{i,L}^2}{2\theta_{i,H}} - \frac{e_{ip,L}^2}{2\theta_{ip,L}}) \end{aligned} \qquad \text{(III.4)}$$

$$\begin{aligned} &q_{ip} \cdot (t_{LH} - \frac{e_{i,L}^2}{2\theta_{i,L}} - \frac{e_{ip,H}^2}{2\theta_{ip,H}}) + (1-q_{ip}) \cdot (t_{LL} - \frac{e_{i,L}^2}{2\theta_{i,L}} - \frac{e_{ip,L}^2}{2\theta_{ip,L}}) \geq \\ &q_{ip} \cdot (t_{HH} - \frac{e_{i,H}^2}{2\theta_{i,L}} - \frac{e_{ip,H}^2}{2\theta_{ip,H}}) + (1-q_{ip}) \cdot (t_{HL} - \frac{e_{i,H}^2}{2\theta_{i,L}} - \frac{e_{ip,L}^2}{2\theta_{ip,L}}) \end{aligned} \qquad \text{(III.5)}$$

We can ignore the IR constraint in (III.3), as this constraint is a combination of constraints (III.4) and (III.1) and hence, the set of feasible contracts in the problem remain exactly the same after (III.3) is dropped. $((t_{HH}, e_{i,H}), (t_{HL}, e_{i,H}) \mid \theta_i = \theta_{i,H}) \geq ((t_{LH}, e_{i,L}), (t_{LL}, e_{i,L}) \mid \theta_i = \theta_{i,H}) > ((t_{LH}, e_{i,L}), (t_{LL}, e_{i,L}) \mid \theta_i = \theta_{i,H}) \geq 0$

The solution to the problem using Kuhn-Tucker conditions follows a similar logic to classic work in economics.[5] The equilibrium effort levels are given by $\hat{e}_{ip,H} = \theta_{ip,H}$, $\hat{e}_{ip,L} = \theta_{ip,L}$, $\hat{e}_{i,H} = \theta_{i,H}$ and $\hat{e}_{i,L} = \theta_{i,L} \cdot \left[\frac{\theta_{i,H} - q_i \cdot \theta_{i,H}}{\theta_{i,H} - q_i \cdot \theta_{i,L}} \right]$ leading effort in state $\theta_{i,L}$ be below both the level of effort when θ_i is observed, $\hat{e}_{i,L} < \theta_{i,L} = e_{i,L}^*$ as well as below the level of effort in state $\theta_{i,H}$, $\hat{e}_{i,L} < \theta_{i,H} = \hat{e}_{i,H}$. For the technical notes regarding the solution above, see Appendix B.

While the reduction in variability does not bring the insurer to a first-best contracting situation, it does lessen the amount of surplus transfer to the practice which would be required under the second-best solution without third-party screening. Put differently, conditioning on a third-party extracted signal enables the insurer to reduce the informational rent and consequently reduce the distortion in effort for type θ_L. When

[5] In particular, work on asymmetric information with hidden information such as Pigou's (1912) second-degree price discrimination representation, the Rothschild & Stiglitz (1976) model of adverse selection in insurance markets, or Maskin & Riley (1984) model of auction design.

the optimal level of effort is chosen, the effort ratio in high and low states is proportional to the realized efficiency parameter. If the signal *s*, accurately reveals the realized value of θ_{ip}, the insurer can eliminate the distortion to effort e_{ip}, while such distortion persists for e_i.

The left hand-side panel of Figure 1 plots the contracts under an observed θ_{ip}, where the transfer level is depicted on the vertical axis and effort level is depicted on the horizontal axis. The blue convex curves represent the practice indifference curves for each realization of θ_{ip}, i.e. all *t* and e_{ip} pairs that makes the practice indifferent between taking the contract and not. These curves start at the origin as, $(t, e_{ip}) = (0,0)$ represents the normalized opportunity cost of effort. A shift of the practice indifference curve to the northwest results in a positive surplus for the practice. The red linear curves represent the insurer iso-profit curves for each realization of θ, i.e. all *t* and e_{ip} pairs that make the insurer maintain a fixed level of profit, π. A shift of the insurer iso-profit curve to the southeast results in an increase in surplus for the insurer. Therefore, the optimal contracts for the insurer, conditional on the practice's participation constraint, will always be the tangency between the two curves. In these points the practice is required to set its effort level to the point at which the marginal benefit exactly equals its marginal cost.

[FIGURE 1]

Note that the first-best solution is no longer attainable as it is clear that type $\theta_{i,H}$ will derive higher surplus from choosing the contract directed at practice type $\theta_{i,L}$ (the dotted line). Hence, the IC constraint for $\theta_{i,H}$ must be binding. The two second-best contracts $(\tilde{t}_H, \tilde{e}_{i,H}) \in \mathbb{R}$ x $\mathbb{R}_+$ for state θ_H and $(\tilde{t}_L, \tilde{e}_{i,L}) \in \mathbb{R}$ x $\mathbb{R}_+$ for

state $\theta_{i,L}$ that solve the problem yield equilibrium effort levels $\tilde{e}_{i,H} = \theta_{i,H}$ and $\tilde{e}_{i,L} = \theta_{i,L} \cdot \left[\frac{\theta_{i,H} - q_i \cdot \theta_{i,H}}{\theta_{i,H} - q_i \cdot \theta_{i,L}} \right] < \theta_{i,H} = \tilde{e}_{i,H}$.

The right hand-side panel of Figure 1 illustrates the equilibrium for θ_i, in which the contract directed at type $\theta_{i,L}$, $(\tilde{t}_L, \tilde{e}_{i,L})$ is no longer incentive compatible (i.e. the practice's indifference curve is not tangent to the iso-profit curve). Hence, a type $\theta_{i,L}$ practice will exert less effort, $\tilde{e}_{i,L} < e^*_{i,L}$ and receive a lower transfer $\tilde{t}_L = t^*_L$, keeping it on the same original indifference curve with reservation utility zero. Practice type $\theta_{i,H}$ will exert the same (efficient) level of effort, $\tilde{e}_{i,H} = e^*_{i,H}$, but receive a higher transfer, $\tilde{t}_H = t^*_H$, placing it on a higher indifference curve with a reservation utility greater than zero. The reason for the transfer of surplus from the insurer to the practice lies in the fact that a $\theta_{i,H}$ type can disguise as a $\theta_{i,L}$ type, which in turn forces the insurer to give up surplus in order to have $\theta_{i,H}$ reveal its true type.

Now consider the case where either both $\theta_{p,L} + \theta_{ip,H}$ and $\theta_{p,H} + \theta_{ip,L}$ belong to s_H or alternatively both $\theta_{p,L} + \theta_{ip,H}$ and $\theta_{p,H} + \theta_{ip,L}$ belong to s_L. In these cases, the insurer cannot use the signal to accurately detect the value of θ_{ip}. Nevertheless, the signal may still be desired as its value is not zero. If both $\theta_{p,L} + \theta_{ip,H}$ and $\theta_{p,H} + \theta_{ip,L}$ belong to s_L, $\theta_{ip} = \theta_{ip,H}$ is detected when $s = s_H$. Similarly, if both $\theta_{p,L} + \theta_{ip,H}$ and $\theta_{p,H} + \theta_{ip,L}$ belong to s_H, $\theta_{ip} = \theta_{ip,L}$ is detected when $s = s_L$. Hence, two elements raise the value of quality assurance recognition to the insurer – the strength of the signal in detecting the true value of θ_{ip}, and the contribution of e_{ip} (relative to e_i) to the insurer's surplus. In other words, the value of the signal is greater the larger the scope for actions that benefit both patients and the insurer and the more

accurate the NCQA signal is in detecting how efficient the practice is in taking these actions, and the more profitable it is for the insurer to ensure that the practice exerts the utmost effort in that regard.

This model ignores an important aspect of the problem – namely, practices have reasons to refuse screening by the credentialing organization. This could happen for two key reasons: the first, the credentialing process may claim resources from the practice, and the second, contracting based on the credentialing signal competes away informational rents that otherwise raise the practices' expected utility. This may explain why in practice, insurance companies have paid the implementation costs associated with implementing the PCMH model, as well as tied the practice per-member per-month reimbursement to the level of NCQA recognition.

IV. *Data*

Program Description

This study analyzes data from HMO enrollees in a single large, commercial insurer in Pennsylvania. 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.[6] 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.

[6] 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.

Recognition Data

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.

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, 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. Figure 2 represents the cumulative number of PCMH-recognized practices over time. The top blue line tracks the timing of recognition for all 280 PCMH in this market between March 2008 and September 2012. The orange line below it tracks the 225 adult PCMHs (excluding 55 pediatric practices). The green line below it tracks the 134 practices for which detailed NCQA recognition data was obtained, and finally, the bottom red 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. .

[Figure 2]

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.[7] Applying these exclusion criteria generated two datasets: the first includes 370,764 patient-year observations; and a second, referred to as "patient panel sample," is comprised of 105,606 patient-years for members enrolled in the same practice during all 6 years of the study period. The patient 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. For robustness, we conducted separate sub-analyses using only chronically ill patients in order to ascertain whether and to what extent the transition to a PCMH model affects such patients differentially.

[7] 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.

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.

[TABLE 2]

Several features of Table 2 warrant further discussion. First, note that patients in the "panel" sample (those with 6 years of continuous enrollment in the data) are somewhat older than in the "full" sample, and that patients with chronic illnesses are substantially older than either group. 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. We also observe some variability in age across clusters, with Cluster 1 serving substantially younger patients than 2 or 3.

Additionally, along with this difference in age across clusters, we see a substantial difference in terms of patient health. The first panel of Table 2 points to patients in Cluster 3 practices having the highest risk scores, followed by Clusters 2 and 1. These differences are likely attributable to the age differences across clusters as well as the modest differences in comorbidity prevalence presented in Table 2, but as the gradient across clusters persists in the panel sample, it is unlikely for these differences to reflect selective entry or attrition of patients.

Patient Outcomes Data

A summary table of patient outcomes, overall and by cluster, appears in the bottom panel of Table 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.[8] 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).

Of particular importance is the long time series analyzed in this study, allowing exploration of the long-run effects of the PCMH on patient outcomes. The earliest practice recognitions in our data occur in 2008, so even with the exclusion of the transition year, we are able to observe up to five years of data post-recognition for some patients. This represents a significant advantage over many previous studies with more limited follow-up periods.

IV. Empirical Approach

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, precluding 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

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

(Shortell, Wu, Lewis, Colla, & Fisher, 2014). The 2008 NCQA recognition process includes 139 unique components.[9] 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.[10]

We employed an agglomerative hierarchical clustering approach to group similar practices.[11] This approach begins with *N* clusters of size one, and sequentially combines the most similar practices into larger clusters. Similarity of practices was measured using the Jaccard distance, and practices were grouped using Ward's method, which minimizes within-cluster variance. This process produces a vector of possible clustering solutions, ranging from the starting point (*N* clusters of size one) to a single cluster of *N* units. For the analysis which follows, a single clustering solution was chosen through inspection of a dendrogram, which provides a visual representation of the increase in within-cluster distance after each aggregation. Ultimately, a three-cluster solution was chosen.[12] Figure 3 shows a plot of the dendrogram used in this process.

[Figure 3]

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 each one of the 30 mean element-level scores (that is, the level of aggregation above factors) differ by cluster.

[9] 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. 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.

[10] 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).

[11] The commonly-used *k-means* clustering approach may have undesirable properties when attempting to identify clusters in higher-dimensional space (Sun, Wang, & Fang, 2012).

[12] The dendrogram in Figure 1 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.

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. 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.

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 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.

However, a regression approach which estimates these effects using a single indicator for the post-treatment period (as in David et al. (2014)) 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:

$$Y_{ijt} = \lambda_t + \mu_{i/j} + \beta X_{ijt} + \sum_{c=1}^{C} \delta_c \, (Post\ x\ Cluster_c)_{it} + \varepsilon_{ijt} \qquad \text{(IV.1)}$$

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), (2) the number of such encounters, conditional on having at least one, or (3) 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.

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).[13] λ_t is a year fixed effects term, and $\mu_{i/j}$ represents practice or 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 (1) 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.

Robustness Analyses

The features of the age distribution and rates of chronic illness across clusters may have significant implications for our analyses. First, it may be the case that practices are choosing which elements of the PCMH model to implement based on idiosyncratic practice characteristics such as their specialty type or patient mix.

Additionally, there has been debate in the literature about whether the PCMH model should be applied to the treatment of all patients, or targeted to patients with chronic illnesses (Berenson et al., 2008). Prior analyses have indicated that the PCMH model may differentially benefit patients with chronic illnesses (David et al., 2015). In light of these considerations, we restrict our sample to patients identified as having chronic illnesses. While our results are not affected by excluding patient risk score as an explanatory

[13] 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.

variable, we explore whether PCMH adoption is affecting patient risk score, by repeating the same specifications (excluding the risk score) and treating the score as a dependent variable. If the risk score is reduced by implementation of the PCMH due to improvements in care, our estimates may be biased in models which include this as a control.

V. Results

Clustering Results

As discussed above, the dendrogram in Figure 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 4.

[FIGURE 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 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 35.5%, and 33.3%, respectively. Another example is element B2 – *Electronic system for clinical data* – which received an average score of 100% and 98% in Clusters 1 and 3, while receiving an average score of only 41.7% in cluster 2. Put differently, nearly all practices in Clusters 1 and 3 have fully implemented B2, while for practices in Cluster 2 there is partial implementation. The score for Cluster 2 can reflect a situation where only two in five 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 1 and 3, which represents the case where the Clusters 1 and 3 dominate Cluster 2. A complete summary of the ANOVA analysis appears in Appendix Table A1.

While most elements are easily assignable to a given location on the diagram, element E3 – *Continuity of care* - did 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 87% while Cluster 2 received an average of 52%. 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.

[APPENDIX TABLE A1]

In Figure 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 1 and 3, based on the labels highlighted in Figure 4. Clusters 1 and 3 dominate Cluster 2 in terms of average total points received – 82 and 83.7 versus 55.9. Hence, the relatively few factors recorded for Cluster 2 are also shared by Clusters 1 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 1 and 3 (presented in areas that belong to a single cluster). Cluster 1 appears to be focusing in population management activities, whereas Cluster 3 emphasizes implementation of decision support, data reporting, and enhanced access.

A second Venn diagram, expressing the differences in emphasis of the 30 elements by NCQA recognition level, appears in the Appendix as Figure A1.

[APPENDIX FIGURE A1]

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.[14] 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.

Some of the elements have clear connections to the outcomes of interest. For example, Cluster 1 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

[14] 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.

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 1 and 3 to dominate Cluster 2, which has both a lower average overall practice score and no areas of implementation where it has clear advantages over Clusters 1 or 3. We can make some further predictions about relative performance in terms of the specific areas where Clusters 1 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.

Patient Outcome Regression Results

Primary Care and Specialist Visit Outcomes

The tables which follow present results from a series of specifications for each outcome of interest. The top row in each table specifies the outcome of interest. In the following row, we find three specification labels – PCP fixed effects, member fixed effects, and member fixed effects (panel) – indicating that the $\mu_{i/j}$ term in the regression specification references to physician or patient fixed effects. In the third case, the sample is restricted to the panel of continuously enrolled patients. The third row repeats two labels under each specification heading: "all adult" and "chronic," indicating whether the model is estimated using all adult patients, or the subset of adult patients with one of the chronic illnesses of interest. The coefficient estimates presented in the table represent percentage point changes in the case of binary outcomes, fractional visits in the case of the conditional utilization outcomes, and dollar changes in the expenditure tables. In the descriptions of the results which follow, the coefficient estimates from the tables are presented as a percent

increase or decrease relative to overall rates for the outcomes in question. 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.

For the Panel A results the baselines used are the outcome rates in the full sample and 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.

[TABLE 3]

Panel A of Table 3 displays the difference-in-differences coefficients for the single PCMH adoption indicator on the probability of having one or more primary care or specialist visits in a given year. These results provide only limited evidence for an overall PCMH effect on PCP utilization, with a significant estimate in only one specification. We find more consistent evidence of a roughly 2% increase in specialist visits among the chronically ill subsample, though these estimates are only marginally significant in the PCP fixed effects model, and non-significant in the stable panel. In Panel B, we find similarly weak evidence for any effect of the PCMH levels – only one specification yielded a significant increase in PCP visits, along with two marginally significant results for specialist utilization.

Turning to Panel C, we find the results from the specification using separate difference-in-differences terms for each of the four 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. With respect to primary care visits, we find consistent evidence of an increase in utilization in Clusters 1 and 3 (2.8-4.4% and 1.9-4.1%, respectively), and a large decrease in Cluster 2 (8.0-13.1%). These results are robust to the use of patient or practice fixed effects, the full sample or the chronically ill only as well as to limiting the analysis to the stable panel. Turning to the specialist results, we find evidence of increased specialist utilization among the chronically ill subsample in Cluster 1 (2.8-

5.0%), and decreased utilization in the full sample in Cluster 3 (2.3-3.4%). Only one of the models produced a significant estimate for Cluster 2, and the direction of the effect is inconsistent across specifications.

[TABLE 4]

Table 4 presents the estimated effects on primary care and specialist visits with respect to intensive margins (the number of visits, conditional on having at least one in a given year). Again, we find limited and mixed evidence of an overall PCMH effect on utilization. We find no significant effects on the number of specialist visits, whether using the single PCMH indicator (Panel A) or level-specific indicators (Panel B). With respect to PCP visits, we find two marginally significant effects in the chronic samples (though with conflicting signs), and a less ambiguously significant decrease in the all adult panel sample, with the results reported in panel A. In Panel B, we find evidence that these results are driven by an apparent decrease in the conditional number of visits among the stable panel of patients in Level 1 and 2 practices, and an increase in visits in Level 3 practices, though the sign and significance vary across specifications.

However, in Panel C, we find results for the individual cluster effects on the number of PCP visits which largely echo those in Table 3. We find significant increases in the number of visits in Clusters 1 and 3, except in the patient panel specifications, and across the board decreases in the number of visits in Cluster 2. On the other hand, we find only one significant effect on the number of specialist visits, with inconsistent signs across specifications.

Inpatient and Emergency Department Utilization

Table 5 displays the output from regressions predicting the probability of having at least one inpatient hospitalization or ED visit in a given year. In Panel A, we find two specifications produced significant estimates of the overall PCMH effect on the probability of being hospitalized, predicting an increased chance of being admitted, when patient fixed effects are employed. With respect to ED utilization, we find

only one (marginally) significant effect of the PCMH as a whole. In the results from the specifications using the level terms reported in Panel B, we similarly find generally marginally- or non-significant results.

[TABLE 5]

In Panel C, we find additional evidence of a PCMH effect by cluster, though the results are again mixed. We observe significant increases in hospitalizations in the two patient fixed effects specifications in the chronically ill subsample for Cluster 1, and significant decreases in the full sample for Cluster 3 – though not in the stable panel or in the chronically ill subsamples.

For ED utilization, we find no significant effects on the probability of one or more visits in Clusters 1 or 3, but some evidence of increased ED utilization in Cluster 2 (ranging from 8.1-14.8% among the significant results, and 4.7-14.8% overall).

[TABLE 6]

The results for the conditional hospitalization and ED visit regressions appear in Table 6. Regarding the overall PCMH results for number of hospitalizations in Panel A, only the estimates using the stable patient panel and patient fixed effects are significant, though the sign is consistently negative across all six estimates. These results are echoed in Panel B, though with reduced significance. In Panel C, we again find that the estimates of the effect of PCMH adoption are confined to the panel sample, and limited to Clusters 1 and 2. In all three cases, the estimate is smaller in magnitude and only marginally significant when analyzing all patients (as opposed to the chronically ill subsample). We find no evidence of effect of the overall PCMH on ED utilization, and only two marginally significant estimates when looking at the clusters individually.

Expenditures for Physician Services

Table 7 presents a final set of outcomes – expenditures on professional services.

[TABLE 7]

Consistent with the previous tables, we find no overall effect of the PCMH model, or when using the level-specific interaction terms, on expenditures for professional services. However, in Panel C we observe a consistent PCMH effect across all six specifications for Cluster 3, with reductions in expenditures of between 4.2 and 18.0% relative to baseline.

Specification Check – Risk Score Regressions

Finally, the right side of Table 7 presents results from a set of regressions treating the risk score as an outcome of interest. Successful implementation of the PCMH model may reduce the burden of disease for the clinic members and represent a benefit that extends beyond short run monetary savings or reduction in utilization. In the non-panel specifications, it may capture selection issues, either riskier patients who select into a PCMH or clinics cherry picking patients. In addition, this serves as a specification check to assess whether adoption of the PCMH model impacts the risk score used as a patient control in all of the models presented above. If the medical home's chronic illness management components were sufficiently effective, it is possible that adoption might reduce the apparent severity of illness and consequently bias our estimates of the PCMH effect when controlling for risk. However, we find no such effect, either of the PCMH overall or the individual clusters (Panels A and C), suggesting that this potential source of bias is not confounding our results. The two "all adults" non-panel specifications returned significant reductions in risk score for practices with Level 2 certification in Panel B. While not statistically significant, the "chronic" sample yields increases in risk. This suggests potential favorable selection of members with no chronic conditions in Level 2 practices.

VI. Discussion

Theoretically, primary care reform 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. Our theoretical model has a number of important features which can inform considerations of these topics. First, the practice can allocate resources to different subsets of actions that would benefit stakeholders (namely patients and insurers) differentially. For example, insurers would benefit from primary care providers substituting lower-cost specialty service (like diagnostic imaging) providers, so long as these substitutions do not entail quality reductions that increase spending downstream. Alternatively, given a fixed threshold for recognition, providers will likely choose to achieve PCMH status by implementing the subset of actions where they find the lowest costs of improvement, rather than the ones most likely to (for example) reduce total spending. Understanding these dynamics can help explain why the incentives put in place by the PCMH recognition process do not generally reduce costs.

Further, the type of signal which NCQA recognition provides – indexing the different actions into a single metric – makes it difficult to fully overcome practice-level adverse selection; nonetheless, insurers (even if their motives are pure profit maximization) will likely find value in credentialing signals. To be clear, insurers could benefit much more from observing the different capacities for actions at the practice level. By collapsing the various dimensions of capability into a single practice score (here, the PCMH recognition level), the NCQA limits the ability of insurers to contract on them. However, the size of the efforts which benefit both patients and insurers, (e_{ip}), is crucial. If it is zero, NCQA recognition is meaningless for a pure profit maximizing insurer, as the signal provided will reflect only gains which accrue to patients. If, on the other hand, the scope of activities which potentially benefit both patients and insurers, (e_{ip}), is quite broad, insurers could benefit significantly from even a noisy signal. The PCMH model's focus on areas like decision support, data reporting and access to care is consistent with trying to provide such a signal - both parties stand to gain from avoiding utilization of intensive medical services that is likely to follow from lack of access to primary care and breaks in patient-physician communication.

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 empirical results, 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, and that 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 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 (1 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.

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. As highlighted in the theoretical model in Section III, practices may enjoy informational rents when the specifics of implementation are not revealed to the insurer.

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 V.1. We do find some evidence of reduced PCP contacts in Cluster 1, consistent with the advanced communications tools, though this does not translate into significant reductions in spending. Table 5 also presents some evidence of an increase in ED utilization in Cluster 1, which conflicts with our priors given the emphasis on population management. Despite dominating on overall PCMH points and with respect to implementation of specific PCMH components, Clusters 1 and 3 do not (generally) dominate Cluster 2 in terms of post-implementation improvements in hospitalizations or ED utilization, though we do find reductions in professional spending for Cluster 3.

Finally, considering the results using only the restricted sample of patients with chronic illness, we would expect to find larger and more significant effects of PCMH implementation, because these patients should stand to gain the most from the introduction of improved population health management. However, we find little evidence of this being the case in our sample.

VII. 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.

References

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

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. http://doi.org/10.1377/hlthaff.27.5.1219

Bodenheimer, T., & Pham, H. H. (2010). Primary care: Current problems and proposed solutions. *Health Affairs*, *29*(5), 799–805. http://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.

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

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. http://doi.org/10.1056/NEJMp1013067

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. http://doi.org/10.1111/1475-6773.12218

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

Engel, C. (2006). Corporate Design for Regulability : A Principal-Agent-Supervisor Model. *Journal of Institutional and Theoretical Economics*, *162*(1), 104–124.

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), 815–25. http://doi.org/10.1001/jama.2014.353

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. http://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. http://doi.org/10.1177/1077558712453618

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

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.

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. http://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.

Khalil, F., & Lawarree, J. (1995). Collusive Auditors. *American Economic Review: Papers and Proceedings*, *85*(2), 442–446.

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. http://doi.org/10.1377/hlthaff.2010.0012

Lewis, T. R., & Sappington, D. E. M. (1991). Incentives for monitoring quality. *The RAND Journal of Economics*, *22*(3), 370–384. http://doi.org/10.2307/2601053

Liu, P.-W. (1982). Imperfect Information, Monitoring Cost and Incentives in the Agency Problem. *Journal of Economics*, *42*(2), 235–245.

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

Maskin, E., & Riley, J. (1984). Optimal Auctions with Risk Averse Buyers. *Econometrica*, *52*(6), 1473–1518. http://doi.org/10.2307/1913516

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.

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*. http://doi.org/10.1111/1475-6773.12085

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

Pigou, A. C. (1912). *Wealth and welfare*. Macmillian and Company, limited.

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

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), 835–43. http://doi.org/10.1377/hlthaff.2010.0158

Rothschild, M., & Stiglitz, J. (1976). Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information. *The Quarterly Journal of Economics*, *90*(4), 629. http://doi.org/10.2307/1885326

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*, n/a–n/a. http://doi.org/10.1111/1475-6773.12234

Strausz, R. (1997). Delegation of Monitoring in a Principal-Agent Relationship. *Review of Economic Studies*, *64*(3), 337. http://doi.org/10.2307/2971717

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. http://doi.org/10.1214/12-EJS668

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

Vafaï, K. (2005). Collusion and organization design. *Economica*, *72*(285), 17–37. http://doi.org/10.1111/j.0013-0427.2005.00400.x

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

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

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

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. http://doi.org/10.1377/hlthaff.2009.0748

Zhao, R. R. (2008). All-or-Nothing Monitoring. *The American Economic Review*, *98*(4), 1619–1628.

Table 1: 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 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	105,606	25,265	232,699	82,231	55,834	44,526	17,713	13,081
PATIENT CHARACTERISTICS										
Age	48.45	61.51	54.29	65.23	47.86	49.50	49.36	60.76	63.15	61.84
	(19.26)	(18.00)	(18.81)	(16.53)	(19.09)	(19.71)	(19.19)	(18.18)	(17.72)	(17.57)
Gender (Female)	0.56	0.54	0.57	0.54	0.57	0.55	0.57	0.55	0.51	0.56
	(0.50)	(0.50)	(0.49)	(0.50)	(0.50)	(0.50)	(0.49)	(0.50)	(0.50)	(0.50)
Risk Score	2.76	6.92	2.91	6.27	2.62	2.93	3.11	6.59	7.37	7.43
	(7.35)	(12.89)	(6.80)	(10.99)	(7.01)	(7.68)	(8.19)	(12.43)	(13.15)	(13.99)
Congestive Heart Failure	0.02	0.10	0.02	0.09	0.02	0.02	0.02	0.09	0.10	0.10
	(0.14)	(0.30)	(0.15)	(0.29)	(0.13)	(0.15)	(0.15)	(0.29)	(0.31)	(0.30)
Chronic Obstructive Lung	0.03	0.15	0.03	0.14	0.03	0.03	0.04	0.14	0.15	0.18
Disease	(0.17)	(0.35)	(0.18)	(0.35)	(0.16)	(0.18)	(0.20)	(0.34)	(0.35)	(0.38)
Coronary Artery Disease	0.06	0.30	0.08	0.33	0.06	0.07	0.07	0.29	0.34	0.29
	(0.24)	(0.46)	(0.27)	(0.47)	(0.23)	(0.26)	(0.25)	(0.45)	(0.47)	(0.45)
Asthma	0.07	0.33	0.07	0.29	0.06	0.07	0.08	0.34	0.30	0.34
	(0.25)	(0.47)	(0.25)	(0.45)	(0.24)	(0.25)	(0.27)	(0.47)	(0.46)	(0.47)
Diabetes	0.10	0.48	0.12	0.50	0.09	0.10	0.11	0.47	0.48	0.48
	(0.30)	(0.50)	(0.32)	(0.50)	(0.29)	(0.30)	(0.32)	(0.50)	(0.50)	(0.50)
PRACTICE CHARACTERISTICS										
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)				
UTILIZATION										
Hospitalization	0.08	0.20	0.08	0.18	0.08	0.09	0.10	0.19	0.22	0.22
	(0.28)	(0.40)	(0.28)	(0.39)	(0.27)	(0.29)	(0.29)	(0.39)	(0.42)	(0.41)
PCP Visits	0.81	0.94	0.83	0.94	0.83	0.77	0.81	0.95	0.90	0.95
	(0.39)	(0.24)	(0.37)	(0.23)	(0.38)	(0.42)	(0.39)	(0.21)	(0.30)	(0.22)
Specialist Visits	0.47	0.75	0.51	0.76	0.46	0.47	0.49	0.74	0.76	0.75
	(0.50)	(0.43)	(0.50)	(0.43)	(0.50)	(0.50)	(0.50)	(0.44)	(0.43)	(0.44)
ED Visits	0.13	0.19	0.12	0.17	0.13	0.14	0.14	0.19	0.20	0.19
	(0.34)	(0.39)	(0.33)	(0.37)	(0.34)	(0.34)	(0.35)	(0.39)	(0.40)	(0.39)
Professional	2,085.28	3,723.21	2,082.98	3,410.73	2,043.59	2,170.51	2,134.67	3,645.60	3,852.71	3,812.45
Expenditures	(6,856.16)	(7,416.81)	(5,094.82)	(6,267.11)	(7,440.29)	(5,523.45)	(6,024.49)	(7,110.36)	(7,562.22)	(8,189.90)

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

Table 3: Primary Care and Specialist Visits – Any Visit

	Primary Care Practice Visits						Specialist Visits					
	PCP Fixed-Effects		Member Fixed-Effects		Member FE (Panel)		PCP Fixed-Effects		Member Fixed-Effects		Member FE (Panel)	
	All Adult	Chronic	All Adult	Chronic	All Adult	Chronic	All Adult	Chronic	All Adult	Chronic	All Adult	Chronic
A - Baseline Model												
POST	0.0049	-0.00132	**0.00714****	0.00605	-0.00081	-0.00378	0.00173	**0.0155***	-0.000799	**0.0184****	-0.00481	0.0163
	[0.00327]	[0.00459]	**[0.00329]**	[0.00512]	[0.00491]	[0.00716]	[0.00397]	**[0.00807]**	[0.00393]	**[0.00800]**	[0.00609]	[0.0114]
Observations	370,764	75,320	370,764	75,320	105,606	25,265	370,764	75,320	370,764	75,320	105,606	25,265
R-squared	0.096	0.064	0.623	0.599	0.467	0.443	0.178	0.129	0.669	0.705	0.545	0.589
B - Level x POST interactions												
Level 1 x POST	0.0021	0.0027	0.0024	0.0046	0.00205	0.00355	-0.00404	-0.00199	-0.00105	0.0215	-0.0103	0.0222
	[0.00560]	[0.00765]	[0.00601]	[0.00907]	[0.00811]	[0.0114]	[0.00683]	[0.0136]	[0.00721]	[0.0144]	[0.0101]	[0.0186]
Level 2 x POST	0.0018	0.0024	0.0018	0.0023	0.0021	0.0025	0.0043	0.0031	-0.0032	0.0195	-0.00642	0.0155
	[0.00528]	[0.00713]	[0.00572]	[0.00854]	[0.00773]	[0.0109]	[0.00643]	[0.00627]	[0.00685]	[0.0135]	[0.00966]	[0.0177]
Level 3 x POST	0.0028	0.0027	**0.0102*****	0.0032	0.0015	0.0024	0.00201	**0.0156***	-0.000461	**0.0178****	-0.00392	0.0156
	[0.00333]	[0.00463]	[0.00337]	[0.00519]	[0.00499]	[0.00717]	[0.00406]	**[0.00826]**	[0.00404]	**[0.00824]**	[0.00624]	[0.0117]
Observations	370,764	75,320	370,764	75,320	105,606	25,265	370,764	75,320	370,764	75,320	105,606	25,265
R-squared	0.099	0.078	0.625	0.606	0.472	0.459	0.178	0.129	0.669	0.705	0.545	0.589
C - Cluster x POST interactions												
Cluster 1 x POST	**0.0297*****	**0.0308*****	**0.0292*****	**0.0360*****	**0.0268*****	**0.0308*****	0.00664	**0.0186****	0.00487	**0.0230*****	0.00328	**0.0208***
	[0.00347]	**[0.00486]**	**[0.00353]**	**[0.00547]**	**[0.00520]**	**[0.00756]**	[0.00422]	**[0.00860]**	[0.00422]	**[0.00860]**	[0.00647]	**[0.0122]**
Cluster 2 x POST	**-0.0703*****	**-0.101*****	**-0.0620*****	**-0.0832*****	**-0.0813*****	**-0.0953*****	-0.00133	0.0153	-0.00564	0.0123	**-0.0161****	0.00601
	[0.00416]	**[0.00576]**	**[0.00436]**	**[0.00666]**	**[0.00612]**	**[0.00881]**	[0.00507]	[0.0102]	[0.00521]	[0.0105]	**[0.00761]**	[0.0142]
Cluster 3 x POST	**0.0221*****	**0.0305*****	**0.0283*****	**0.0334*****	**0.0232*****	**0.0179***	**-0.0121****	0.00563	**-0.0168*****	0.00992	**-0.0172****	0.0166
	[0.00449]	**[0.00606]**	**[0.00487]**	**[0.00732]**	**[0.00669]**	**[0.00953]**	**[0.00546]**	[0.0107]	**[0.00582]**	[0.0115]	**[0.00831]**	[0.0153]
Observations	370,764	75,320	370,764	75,320	105,606	25,265	370,764	75,320	370,764	75,320	105,606	25,265
R-squared	0.096	0.064	0.623	0.599	0.467	0.443	0.178	0.129	0.669	0.705	0.545	0.589

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 4: Primary Care and Specialist Visits – Number of Visits, Conditional on at Least One

	Primary Care Practice Visits (Cond.)						Specialist Visits (Cond.)					
	PCP Fixed-Effects		Member Fixed-Effects		Member FE (Panel)		PCP Fixed-Effects		Member Fixed-Effects		Member FE (Panel)	
	All Adult	Chronic	All Adult	Chronic	All Adult	Chronic	All Adult	Chronic	All Adult	Chronic	All Adult	Chronic
A - Baseline Model												
POST	0.0108	0.0832	-0.0119	**0.115***	**-0.116*****	**-0.168***	0.0475	-0.0682	0.067	0.0738	-0.0732	-0.233
	[0.0271]	[0.0748]	[0.0259]	**[0.0691]**	**[0.0396]**	**[0.0952]**	[0.0628]	[0.142]	[0.0654]	[0.131]	[0.0935]	[0.181]
Observations	307,747	71,738	307,747	71,738	90,013	24,189	179,200	57,297	179,200	57,297	55,676	19,550
R-squared	0.184	0.156	0.724	0.758	0.61	0.666	0.222	0.187	0.766	0.792	0.654	0.681
B - Level x POST interactions												
Level 1 x POST	0.0138	0.0145	-0.0113	-0.0208	**-0.234*****	**-0.329****	-0.106	-0.189	-0.00685	-0.0782	-0.23	-0.217
	[0.0467]	[0.126]	[0.0476]	[0.124]	**[0.0660]**	**[0.154]**	[0.106]	[0.237]	[0.118]	[0.232]	[0.151]	[0.284]
Level 2 x POST	0.0733	0.0162	-0.0884	-0.081	**-0.149****	**-0.301***	0.022	0.028	0.021	0.046	0.041	0.012
	[0.0480]	[0.128]	[0.0507]	[0.129]	**[0.0694]**	**[0.160]**	[0.0982]	[0.217]	[0.109]	[0.215]	[0.143]	[0.271]
Level 3 x POST	0.0105	0.0247	0.0277	**0.1408****	0.0113	0.0445	0.041	0.102	0.0561	0.0198	0.057	0.252
	[0.0276]	[0.0763]	[0.0265]	**[0.0709]**	[0.0403]	[0.0971]	[0.0641]	[0.145]	[0.0671]	[0.135]	[0.0956]	[0.185]
Observations	307,747	71,738	307,747	71,738	90,013	24,189	179,200	57,297	179,200	57,297	55,676	19,550
R-squared	0.184	0.156	0.724	0.758	0.61	0.667	0.222	0.187	0.766	0.792	0.654	0.681
C - Cluster x POST interactions												
Cluster 1 x POST	**0.130*****	**0.277*****	**0.119*****	**0.332*****	0.0481	0.0939	0.0462	-0.0921	0.0392	0.0106	-0.101	-0.301
	[0.0287]	**[0.0795]**	**[0.0276]**	**[0.0740]**	[0.0419]	[0.101]	[0.0664]	[0.151]	[0.0698]	[0.141]	[0.0989]	[0.192]
Cluster 2 x POST	**-0.431*****	**-0.629*****	**-0.479*****	**-0.651*****	**-0.676*****	**-1.047*****	0.0968	0.108	0.131	**0.356****	-0.0763	-0.017
	[0.0355]	**[0.0965]**	**[0.0355]**	**[0.0928]**	**[0.0508]**	**[0.120]**	[0.0791]	[0.177]	[0.0855]	**[0.170]**	[0.115]	[0.221]
Cluster 3 x POST	**0.153*****	**0.350*****	**0.121*****	**0.416*****	0.0605	0.182	-0.0182	-0.234	0.0785	-0.151	0.0369	-0.347
	[0.0372]	**[0.0990]**	**[0.0382]**	**[0.0989]**	[0.0539]	[0.127]	[0.0848]	[0.187]	[0.0947]	[0.188]	[0.126]	[0.241]
Observations	307,747	71,738	307,747	71,738	90,013	24,189	179,200	57,297	179,200	57,297	55,676	19,550
R-squared	0.184	0.156	0.724	0.758	0.61	0.666	0.222	0.187	0.766	0.792	0.654	0.681

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 5: Inpatient Admissions and ED Visits – Any Visit

	Inpatient Hospital Admissions						Emergency Department Visits					
	PCP Fixed-Effects		Member Fixed-Effects		Member FE (Panel)		PCP Fixed-Effects		Member Fixed-Effects		Member FE (Panel)	
	All Adult	Chronic	All Adult	Chronic	All Adult	Chronic	All Adult	Chronic	All Adult	Chronic	All Adult	Chronic
A - Baseline Model												
POST	-0.00177	-0.00329	-0.00019	**0.0164****	0.00361	**0.0197***	**0.00523***	0.00592	0.00501	0.00861	0.00714	0.0137
	[0.00204]	[0.00645]	[0.00235]	**[0.00771]**	[0.00359]	**[0.0107]**	**[0.00293]**	[0.00758]	[0.00325]	[0.00870]	[0.00490]	[0.0120]
Observations	370,764	75,320	370,764	75,320	105,606	25,265	370,764	75,320	370,764	75,320	105,606	25,265
R-squared	0.308	0.352	0.621	0.681	0.481	0.557	0.041	0.055	0.515	0.571	0.312	0.396
B - Level x POST interactions												
Level 1 x POST	-0.003	0.00195	-0.00264	0.014	-0.00097	0.0128	**0.00946***	0.0135	0.00119	0.0205	0.0113	0.0276
	[0.00350]	[0.0109]	[0.00431]	[0.0139]	[0.00597]	[0.0174]	**[0.00503]**	[0.0128]	[0.00595]	[0.0157]	[0.00814]	[0.0196]
Level 2 x POST	0.000585	0.00905	0.00296	0.0206	0.00578	0.0258	0.00131	0.0176	0.00157	0.0197	0.00148	0.016
	[0.00330]	[0.0102]	[0.00409]	[0.0131]	[0.00570]	[0.0166]	[0.00474]	[0.0119]	[0.00565]	[0.0147]	[0.00776]	[0.0187]
Level 3 x POST	-0.00194	-0.00554	-0.00032	**0.0160****	0.00388	**0.0198***	0.00376	0.00348	0.00285	0.00565	0.0056	0.0114
	[0.00208]	[0.00660]	[0.00241]	**[0.00794]**	[0.00368]	**[0.0110]**	[0.00299]	[0.00776]	[0.00334]	[0.00896]	[0.00501]	[0.0124]
Observations	370,764	75,320	370,764	75,320	105,606	25,265	370,764	75,320	370,764	75,320	105,606	25,265
R-squared	0.308	0.352	0.621	0.681	0.481	0.557	0.041	0.054	0.515	0.571	0.312	0.396
C - Cluster x POST interactions												
Cluster 1 x POST	-0.00147	-0.00572	-0.00013	**0.0187****	0.0042	**0.0222***	0.00372	0.00064	0.00408	0.00506	0.00813	0.0124
	[0.00216]	[0.00688]	[0.00252]	**[0.00830]**	[0.00382]	**[0.0114]**	[0.00310]	[0.00808]	[0.00348]	[0.00936]	[0.00520]	[0.0128]
Cluster 2 x POST	0.00066	0.00472	0.00323	**0.0193***	0.00626	0.021	**0.0110*****	**0.0201****	**0.0115*****	0.0178	0.00917	0.0195
	[0.00260]	[0.00814]	[0.00311]	**[0.0101]**	[0.00449]	[0.0133]	**[0.00373]**	**[0.00956]**	**[0.00430]**	[0.0114]	[0.00611]	[0.0149]
Cluster 3 x POST	**-0.00643****	-0.00629	**-0.00609***	0.00297	-0.00278	0.00856	0.00247	0.00375	-0.00176	0.00767	0.000221	0.00885
	[0.00280]	[0.00857]	**[0.00348]**	[0.0111]	[0.00490]	[0.0144]	[0.00402]	[0.0101]	[0.00480]	[0.0125]	[0.00668]	[0.0162]
Observations	370,764	75,320	370,764	75,320	105,606	25,265	370,764	75,320	370,764	75,320	105,606	25,265
R-squared	0.308	0.352	0.621	0.681	0.481	0.557	0.041	0.055	0.515	0.571	0.312	0.396

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 6: Inpatient Admissions and ED Visits – Number of Visits

	Inpatient Hospital Admissions (Cond.)						Emergency Department Visits (Cond.)					
	PCP Fixed-Effects		Member Fixed-Effects		Member FE (Panel)		PCP Fixed-Effects		Member Fixed-Effects		Member FE (Panel)	
	All Adult	Chronic	All Adult	Chronic	All Adult	Chronic	All Adult	Chronic	All Adult	Chronic	All Adult	Chronic
A - Baseline Model												
POST	-0.0233	-0.0468	-0.043	-0.118	**-0.174***	**-0.253****	-0.00261	-0.00237	0.0364	0.102	0.0103	0.0316
	[0.0330]	[0.0557]	[0.0761]	[0.103]	**[0.0919]**	**[0.124]**	[0.0242]	[0.0655]	[0.0447]	[0.108]	[0.0543]	[0.110]
Observations	31,287	15,908	31,287	15908	9,068	4,838	53,750	15,513	53,750	15513	14,057	4,645
R-squared	0.367	0.374	0.845	0.851	0.743	0.758	0.04	0.047	0.813	0.825	0.579	0.584
B - Level x POST interactions												
Level 1 x POST	-0.0619	-0.119	-0.115	-0.15	**-0.236***	**-0.237***	0.00134	0.00159	0.0423	0.292	0.0789	0.274
	[0.0564]	[0.0902]	[0.126]	[0.162]	**[0.136]**	**[0.128]**	[0.0407]	[0.109]	[0.0788]	[0.186]	[0.0870]	[0.172]
Level 2 x POST	-0.00155	-0.0475	-0.0455	-0.113	-0.139	-0.261	-0.00531	-0.00665	0.0207	0.0284	0.0258	0.0275
	[0.0490]	[0.0805]	[0.114]	[0.157]	[0.0941]	[0.177]	[0.0473]	[0.1000]	[0.0773]	[0.176]	[0.0844]	[0.166]
Level 3 x POST	-0.0227	-0.0375	-0.0327	-0.114	-0.196	-0.258	-0.00635	-0.00224	0.0124	0.0421	-0.0359	-0.0525
	[0.0338]	[0.0569]	[0.0781]	[0.107]	[0.128]	[0.172]	[0.0933]	[0.0672]	[0.0462]	[0.112]	[0.0557]	[0.113]
Observations	31,287	15,908	31,287	15908	9,068	4,838	53,750	15,513	53,750	15513	14,057	4,645
R-squared	0.367	0.374	0.845	0.851	0.743	0.758	0.04	0.047	0.813	0.826	0.58	0.585
C - Cluster x POST interactions												
Cluster 1 x POST	-0.0247	-0.0412	-0.0497	-0.12	**-0.187***	**-0.264****	-0.00321	0.0239	0.0189	0.0607	-0.022	-0.0171
	[0.0352]	[0.0594]	[0.0812]	[0.111]	**[0.0966]**	**[0.131]**	[0.0259]	[0.0702]	[0.0485]	[0.118]	[0.0579]	[0.118]
Cluster 2 x POST	-0.0177	-0.0589	-0.0434	-0.117	**-0.200***	**-0.298****	0.00508	-0.0166	**0.105***	0.193	**0.121***	0.144
	[0.0410]	[0.0677]	[0.0947]	[0.127]	**[0.109]**	**[0.145]**	[0.0309]	[0.0819]	**[0.0592]**	[0.138]	**[0.0682]**	[0.134]
Cluster 3 x POST	-0.0271	-0.0459	-0.0169	-0.115	-0.0726	-0.1	-0.011	-0.0624	-0.00414	0.0882	-0.0385	-0.00494
	[0.0438]	[0.0721]	[0.106]	[0.145]	[0.120]	[0.166]	[0.0328]	[0.0870]	[0.0644]	[0.156]	[0.0732]	[0.151]
Observations	31,287	15,908	31,287	15908	9,068	4,838	53,750	15,513	53,750	15513	14,057	4,645
R-squared	0.367	0.374	0.845	0.851	0.743	0.758	0.04	0.047	0.813	0.825	0.579	0.584

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 7: Professional Expenditures (conditional on non-zero) and Risk Score

	Professional Expenditures (PCPs and Specialists)						Risk Score ❖					
	PCP Fixed-Effects		Member Fixed-Effects		Member FE (Panel)		PCP Fixed-Effects		Member Fixed-Effects		Member FE (Panel)	
	All Adult	Chronic	All Adult	Chronic	All Adult	Chronic	All Adult	Chronic	All Adult	Chronic	All Adult	Chronic
A - Baseline Model												
POST	-72.29	-22.82	-5.514	1.811	-74.6	-216.2	0.00165	0.0148	-0.0992	-0.00376	-0.0645	0.377
	[58.60]	[123.7]	[42.06]	[109.0]	[59.23]	[141.6]	[0.0567]	[0.227]	[0.0953]	[0.228]	[0.0843]	[0.304]
Observations	322,539	73,845	322,539	73,845	93,973	24,885	370,764	75,320	370,764	75,320	105,606	25,265
R-squared	0.184	0.314	0.84	0.819	0.657	0.714	0.231	0.216	0.698	0.727	0.530	0.560
B - Level x POST interactions												
Level 1 x POST	151.8	129.4	84.18	54.38	127.1	25.13	0.0623	0.449	-0.0754	0.119	-0.0783	0.676
	[99.82]	[208.1]	[76.26]	[194.6]	[97.51]	[229.3]	[0.0973]	[0.383]	[0.101]	[0.411]	[0.140]	[0.494]
Level 2 x POST	23.9	46.8	35.2	48.2	46.86	-16.1	**-0.223****	0.136	**-0.203****	0.252	-0.181	0.66
	[96.48]	[195.6]	[75.05]	[185.4]	[95.38]	[220.3]	**[0.0917]**	[0.357]	**[0.0964]**	[0.387]	[0.134]	[0.472]
Level 3 x POST	-98.31	-72.42	-13.02	-52.35	-83.37	-210.7	0.0244	-0.0524	-0.0886	-0.0528	-0.0471	0.295
	[59.82]	[126.5]	[43.18]	[112.1]	[60.56]	[145.3]	[0.0579]	[0.232]	[0.0569]	[0.235]	[0.0863]	[0.312]
Observations	322,539	73,845	322,539	73,845	93,973	24,885	370,764	75,320	370,764	75,320	105,606	25,265
R-squared	0.183	0.321	0.838	0.823	0.687	0.702	0.231	0.216	0.698	0.727	0.530	0.560
C - Cluster x POST interactions												
Cluster 1 x POST	-59.52	37.93	26.39	19.57	-51.09	-186.7	0.0488	-0.0147	-0.0725	-0.108	-0.0455	0.254
	[62.11]	[131.8]	[45.02]	[117.1]	[62.80]	[150.8]	[0.0601]	[0.242]	[0.0593]	[0.246]	[0.0895]	[0.324]
Cluster 2 x POST	12.65	99.79	-3.445	194.8	-74.17	-53.48	-0.0873	0.191	-0.162	0.216	-0.129	0.618
	[74.91]	[156.0]	[56.05]	[142.6]	[74.20]	[175.5]	[0.0722]	[0.286]	[0.173]	[0.299]	[0.105]	[0.377]
Cluster 3 x POST	**-237.8*****	**-384.7****	**-140.6****	**-376.6****	**-161.7****	**-575.6*****	-0.0443	-0.129	-0.109	0.0497	-0.0343	0.432
	[79.84]	**[163.7]**	**[61.46]**	**[156.0]**	**[79.98]**	**[189.3]**	[0.0779]	[0.302]	[0.0819]	[0.329]	[0.115]	[0.408]
Observations	322,539	73,845	322,539	73,845	93,973	24,885	370,764	75,320	370,764	75,320	105,606	25,265
R-squared	0.184	0.314	0.84	0.819	0.657	0.714	0.231	0.216	0.698	0.727	0.530	0.560

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

❖ Risk score models omit risk score as explanatory variable

Figure 1: The optimal contracts when θ_{ip} is observed θ_i is unobserved by the insurer

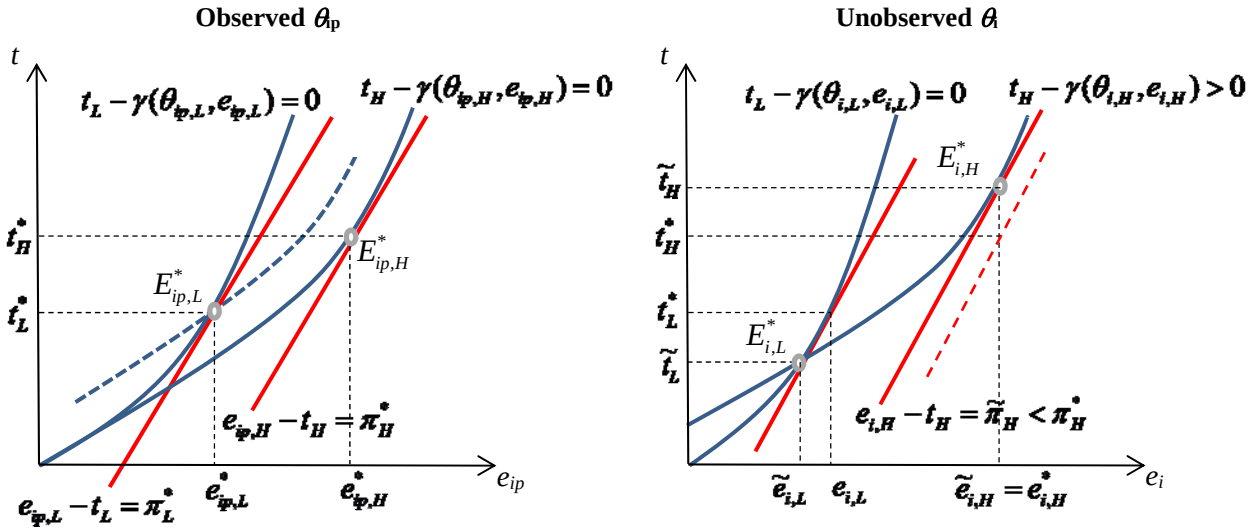


Figure 2

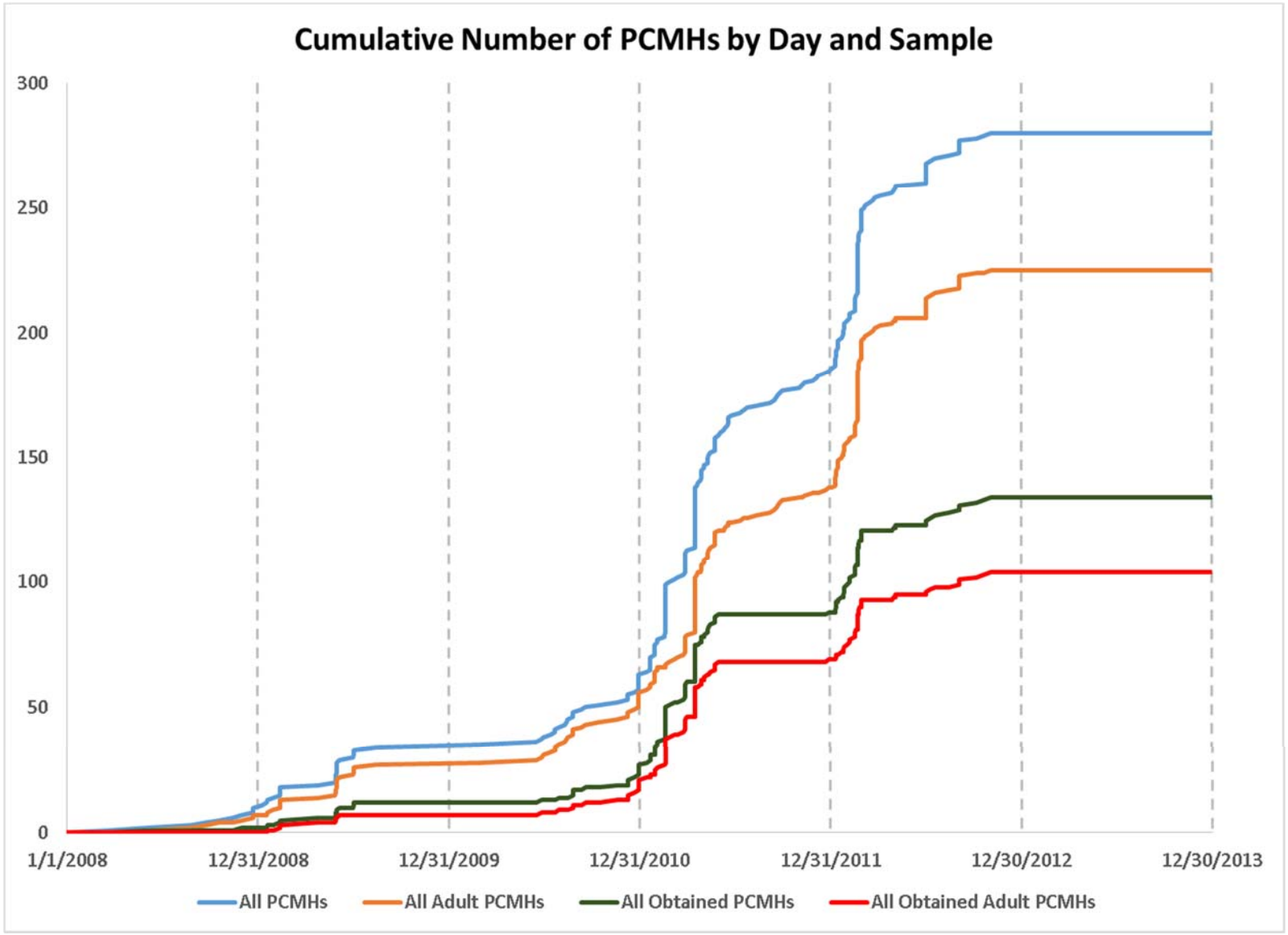

Cumulative Number of PCMHs by Day and Sample
300
250
200
150
100
50
0
1/1/2008
12/31/2008
12/31/2009
12/31/2010
12/31/2011
12/30/2012
12/30/2013
All PCMHs
All Adult PCMHs
All Obtained PCMHs
All Obtained Adult PCMHs

Figure 3: Dendrogram for Hierarchical Clustering Approach

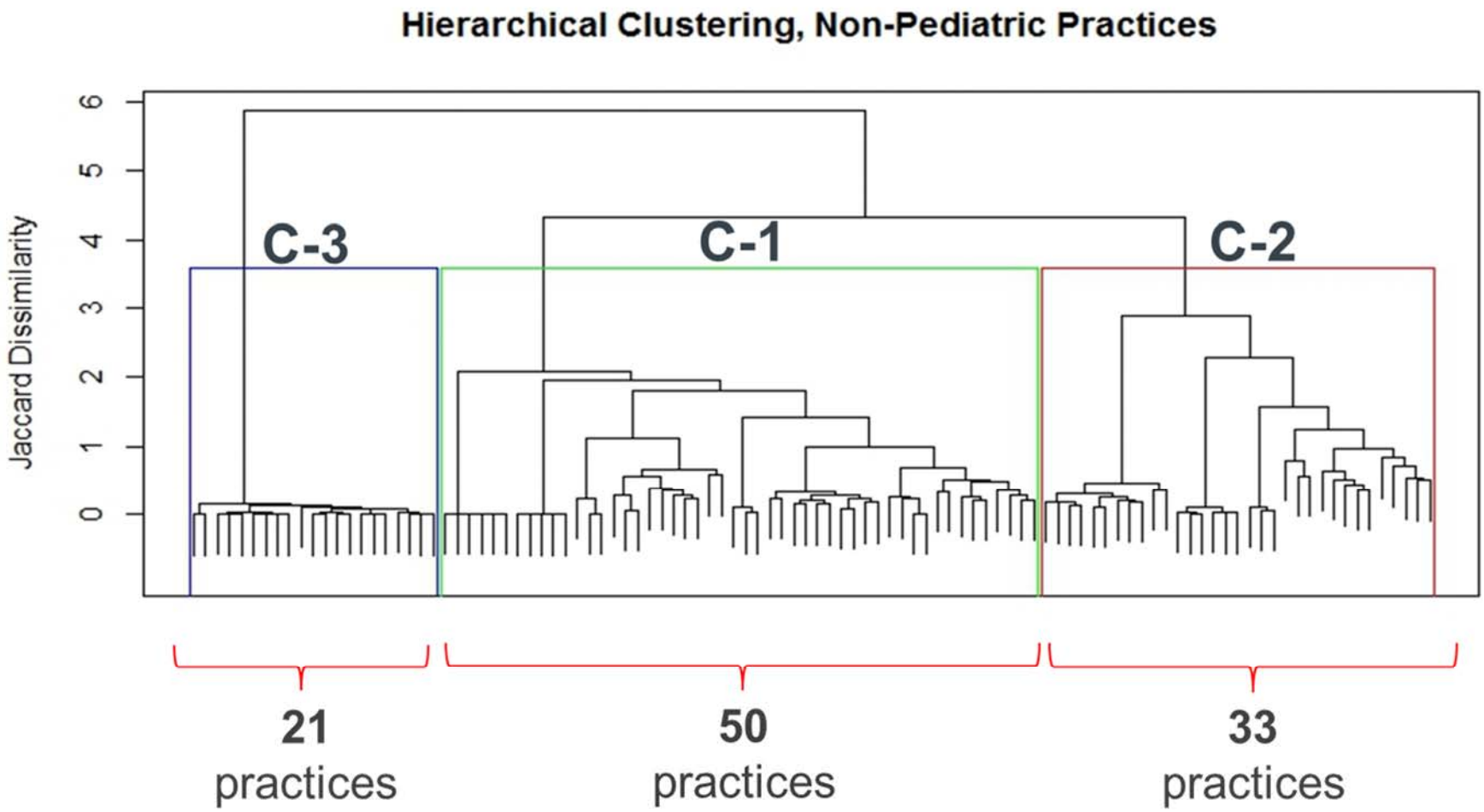
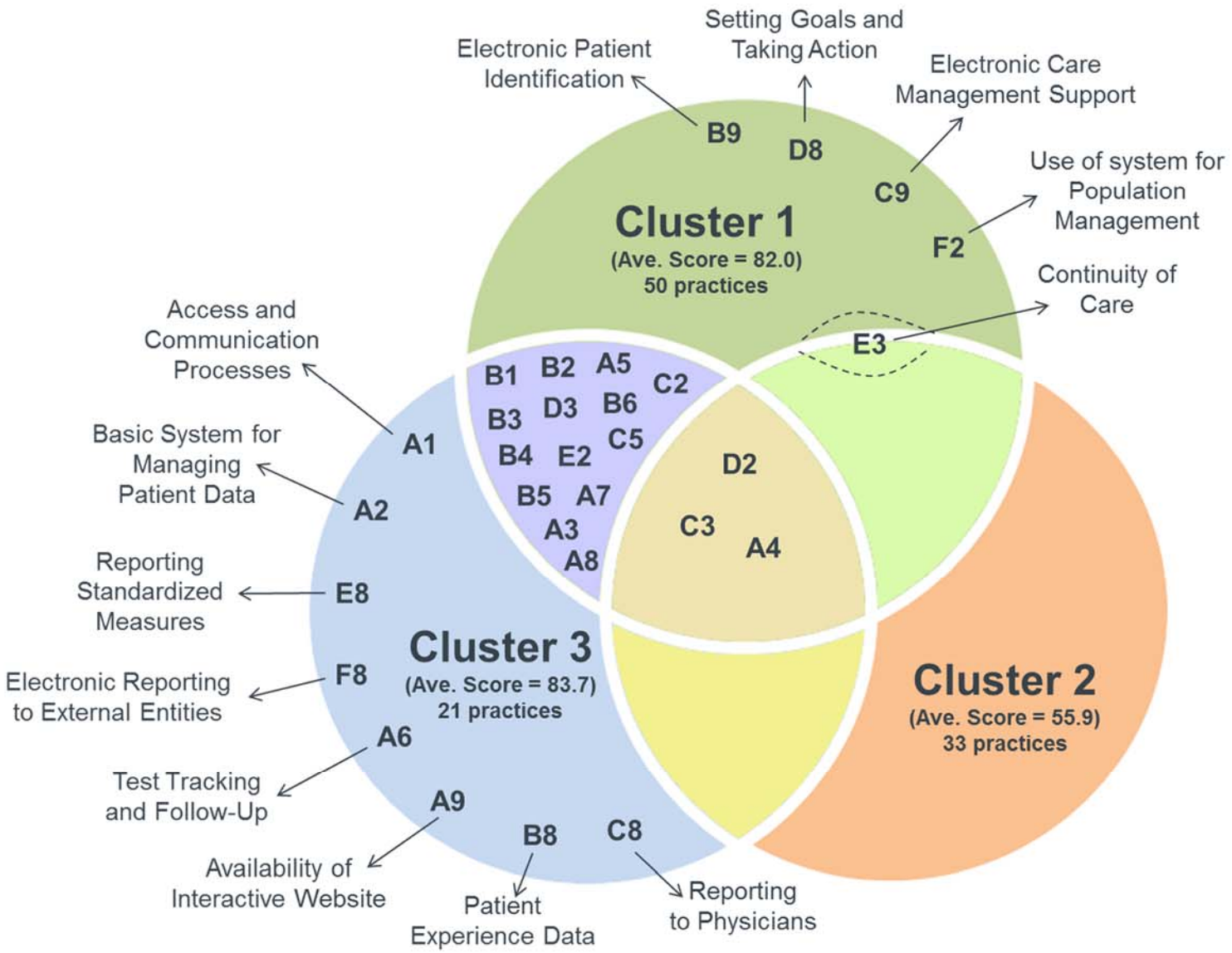
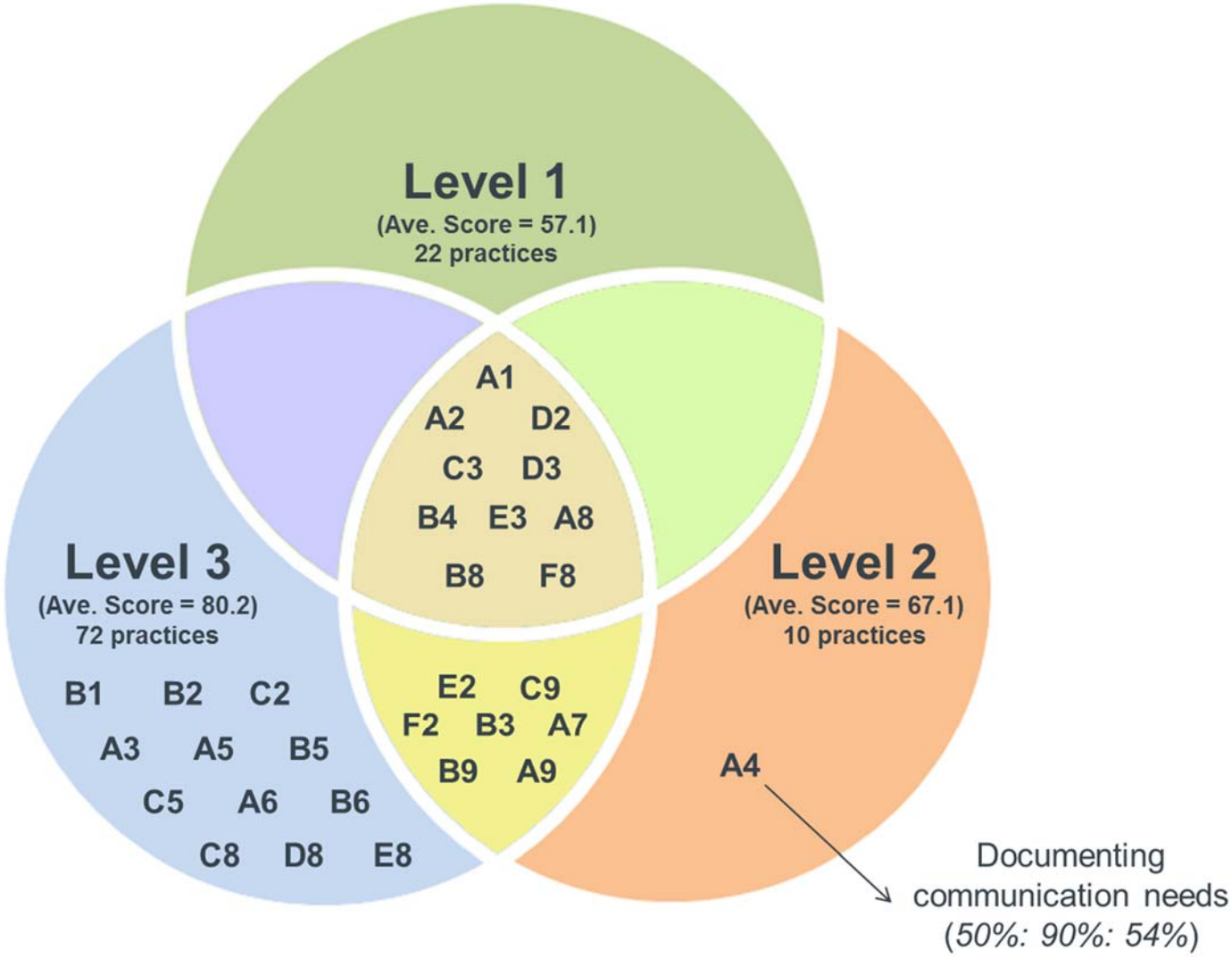


Figure 4: Cluster Description Venn Diagram



Appendix Figure 1: Level Description Venn Diagram



Appendix Table A1: ANOVA Results for Cluster Comparisons

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

Appendix B:

Below is a technical note on the solution of the profit maximization problem is given by:

$$\underset{t_{HH},t_{HL},t_{LH},t_{LL}}{Max} \quad \begin{aligned} E\pi = {} & q_i \cdot q_{ip} \cdot (e_{i,H} + e_{ip,H} - t_{HH}) + q_i \cdot (1-q_{ip}) \cdot (e_{i,H} + e_{ip,L} - t_{HL}) + \\ & + (1-q_i) \cdot q_{ip} \cdot (e_{i,L} + e_{ip,H} - t_{LH}) + (1-q_i) \cdot (1-q_{ip}) \cdot (e_{i,L} + e_{ip,L} - t_{LL}) \end{aligned} \tag{III.4.1}$$

s.t. $$q_{ip} \cdot (t_{LH} - \frac{e_{i,L}^2}{2\theta_{i,L}} - \frac{e_{ip,H}^2}{2\theta_{ip,H}}) + (1-q_{ip}) \cdot (t_{LL} - \frac{e_{i,L}^2}{2\theta_{i,L}} - \frac{e_{ip,L}^2}{2\theta_{ip,L}}) \geq 0 \tag{III.4.2}$$

$$\begin{aligned} & q_{ip} \cdot (t_{HH} - \frac{e_{i,H}^2}{2\theta_{i,H}} - \frac{e_{ip,H}^2}{2\theta_{ip,H}}) + (1-q_{ip}) \cdot (t_{HL} - \frac{e_{i,H}^2}{2\theta_{i,H}} - \frac{e_{ip,L}^2}{2\theta_{ip,L}}) \geq \\ & q_{ip} \cdot (t_{LH} - \frac{e_{i,L}^2}{2\theta_{i,H}} - \frac{e_{ip,H}^2}{2\theta_{ip,H}}) + (1-q_{ip}) \cdot (t_{LL} - \frac{e_{i,L}^2}{2\theta_{i,H}} - \frac{e_{ip,L}^2}{2\theta_{ip,L}}) \end{aligned} \tag{III.4.4}$$

$$\begin{aligned} & q_{ip} \cdot (t_{LH} - \frac{e_{i,L}^2}{2\theta_{i,L}} - \frac{e_{ip,H}^2}{2\theta_{ip,H}}) + (1-q_{ip}) \cdot (t_{LL} - \frac{e_{i,L}^2}{2\theta_{i,L}} - \frac{e_{ip,L}^2}{2\theta_{ip,L}}) \geq \\ & q_{ip} \cdot (t_{HH} - \frac{e_{i,H}^2}{2\theta_{i,L}} - \frac{e_{ip,H}^2}{2\theta_{ip,H}}) + (1-q_{ip}) \cdot (t_{HL} - \frac{e_{i,H}^2}{2\theta_{i,L}} - \frac{e_{ip,L}^2}{2\theta_{ip,L}}) \end{aligned} \tag{III.4.5}$$

Note that constraint (III.4.3) is omitted per our discussion on page 17. Let ψ be the multiplier on the IR constraint (III.4.1) and let λ_1 and λ_2 be the multipliers on the IC constraints (III.4.4) and (III.4.5), respectively. The Kuhn-Tucker conditions for this problem can be written as follows:

Condition with respect to $e_{i,H}$: $$q_i - \lambda_1 \cdot \frac{e_{i,H}}{\theta_{i,H}} + \lambda_2 \cdot \frac{e_{i,H}}{\theta_{i,L}} \leq 0 \tag{A.1}$$

Condition with respect to $e_{i,L}$: $$(1-q_i) - \psi \cdot \frac{e_{i,L:}}{\theta_{i,L}} - \lambda_2 \cdot \frac{e_{i,L}}{\theta_{i,L}} + \lambda_1 \cdot \frac{e_{i,L}}{\theta_{i,H}} \leq 0 \tag{A.2}$$

Condition with respect to $e_{ip,H}$: $$q_{ip} - \psi \cdot q_{ip} \cdot \frac{e_{ip,H}}{\theta_{ip,H}} \leq 0 \tag{A.3}$$

Condition with respect to $e_{ip,L}$: $$(1-q_{ip}) - \psi \cdot (1-q_{ip}) \cdot \frac{e_{ip,L}}{\theta_{ip,L}} \leq 0 \tag{A.4}$$

To insure that all transfer values are positive (i.e. $t_{HH} \geq 0, t_{HL} \geq 0, t_{LH} \geq 0,$ and $t_{LL} \geq 0$), consider the following conditions:

Condition with respect to t_{HH} or t_{HL}: $\quad -q_i + \lambda_1 - \lambda_2 = 0$ (A.5)

Condition with respect to t_{LH} or t_{LL}: $\quad -(1-q_i) + \psi - \lambda_1 + \lambda_2 = 0$ (A.6)

Note that condition (A.5) implies that $\lambda_1 > 0$. Thus, constraint (III.4.4) must hold with equality at the optimal solution. Combining conditions (A.5) and (A.6) implies that $\psi = 1$. Plugging $\psi = 1$ in (A.3) and (A.4) provides the first result: $\hat{e}_{ip,H} = \theta_{ip,H}$ and $\hat{e}_{ip,L} = \theta_{ip,L}$, suggesting that these effort levels are at their optimal levels.

Next, note that both $e_{i,H}$ and $e_{i,L}$ are strictly positive. (A.1) cannot hold if $e_{i,H} = 0$ and (A.2) cannot hold when $e_{i,L} = 0$ as $q_i \neq 0$. We can use (A.1) and (A.5) in conjunction to the information we have uncovered thus far to conclude that $\lambda_1 = q_i$ and $\lambda_2 = 0$. Plugging this solution back in (A.1) and (A.2) we obtain:

$$q_i - q_i \cdot \frac{e_{i,H}}{\theta_{i,H}} = 0 \qquad (A.1^*)$$

$$(1-q_i) - \frac{e_{i,L:}}{\theta_{i,L}} + q_i \cdot \frac{e_{i,L}}{\theta_{i,H}} = 0 \qquad (A.2^*)$$

Rearranging conditions (A.1*) and (A.2*) we receive the second set of equilibrium effort levels: $\hat{e}_{i,H} = \theta_{i,H}$ and $\hat{e}_{i,L} = \theta_{i,L} \cdot \left[\frac{\theta_{i,H} - q_i \cdot \theta_{i,H}}{\theta_{i,H} - q_i \cdot \theta_{i,L}} \right]$.