Information Disclosure and Patient Demand[1]

Nicole Holz

Northwestern University

September 11, 2022

[Click here for latest draft. Use password JMP]

The information contained in this paper is preliminary and incomplete. It is being shared only for the purposes of seeking feedback, input and discussion to help evaluate the study results before the research is deemed final for publication. Please do not cite or circulate this research.

Abstract

Information disclosure programs can help consumers make better choices, but the consumers who respond the most to the information may not benefit the most or generate the most savings for firms designing the programs. I examine a disclosure program used by a private health insurer that highlights ratings of physicians based on two dimensions: quality and cost efficiency. Using a regression discontinuity design, I find that the highest physician ranking of "Premium Care" leads to 38% more new patients (when compared to the lowest ranking) and that the effects are stronger for younger patients. These young patients, however, may not benefit from being matched to higher quality, more cost-efficient physicians. Using a two-way fixed effects research design studying patients who switch physicians following physician exit (from retiring or moving), I find that switching to higher quality, more cost-efficient physicians leads to larger declines in spending for middle-aged patients than for younger adult patients, with no evidence of adverse effects on patient health. Collectively, these results indicate that targeting disclosure programs to middle-aged patients can achieve greater cost savings than web-based ratings systems that disproportionately steer younger patients.

[1] Nicoleholz2023@u.northwestern.edu. Financial support for this research is from the National Science Foundation Graduate Research Fellowship under Grant NSF DGE-1842165, and from the Susan Schmidt Bies Award. I thank David Dranove, Dean Karlan, Matthew Notowidigdo, Molly Schnell, Amanda Starc, and Applied Micro Lunch participants for their guidance and helpful comments. Any opinions expressed in this paper are those of the author, and do not reflect the position of Northwestern University. This paper discusses quality and cost efficiency metrics of physicians, however, all physicians in the network of this insurer have met minimum credentialing requirements which are separate from the qualifications considered by the disclosure program. Just because a given physician does not have the Premium or Quality Care status does not mean that the physician is not a high-quality physician.

1. Introduction

Consumers respond to disclosure reports in many contexts, including education, charity, and healthcare (Allende et al 2019, Dranove and Jin 2010, Mayo 2021,). These disclosure reports are essential in the healthcare context, where spending is high and physicians vary in ways that impact spending and health outcomes (Currie et al 2016, Fadlon and Van Parys 2020). Further, outside of disclosure reports, patients do not have information to assess prospective physicians, especially on characteristics that they may not be able to evaluate or learn from their peers. Disclosure reports aim to solve this problem by disclosing relevant information about physicians. However, the efficacy of such a disclosure program depends on whether the patients who benefit the most or generate the most savings from seeing these physicians are also those who are most likely to respond to the disclosure program. This is especially true in contexts where capacity is constrained since the highest-rated physicians cannot treat all patients. If older patients have more to gain from seeing higher-quality, more cost-efficient physicians, then disclosure policies should target information toward those patients.

I explore these heterogeneous effects in the context of a private health insurer's physician information disclosure program. The insurer creates quality and cost efficiency scores, which are continuous measures, based on compliance with medical practice guidelines (for quality) and payments made by both the insurer and the patients (for cost efficiency). The insurer then uses the scores to assign physicians to one of three possible statuses: "Not Designated" (those whose quality score has not met the quality threshold), "Quality Care" (those who pass the quality threshold, but not the cost-efficiency threshold), and "Premium Care" (those who are both high-quality and cost-efficient) using threshold rules.

The threshold rule facilitates a regression discontinuity (RD) design[2], making three comparisons. First, I explore the comparison of Not Designated physicians to Premium physicians by narrowing in on those who are cost-efficient and studying outcomes around the quality threshold. These cost-efficient physicians who barely pass the quality threshold are given Premium status. My findings show large gains in new patient volume for physicians with a Premium Care status compared to Not Designated: they have 38% more new patient visits due to their designation. The second comparison subsets to non-cost-efficient physicians, who are given a Quality Care

[2] The use of a regression discontinuity design is common in the quality disclosure field. For example, Anderson and Magruder (2012) and Chartock (2021) both use regression discontinuity design to identify demand effects based on online information.

designation if they pass the quality threshold, and Not Designated status if they miss it. These non-cost-efficient physicians experience low gains in passing the threshold. Finally, I compare Quality to Premium Care physicians, using the cost efficiency threshold. Among physicians who meet the quality criteria, those who just pass the cost efficiency threshold are given the Premium designation, and those who just miss it are given the Quality designation. The Premium designation does not lead to significant gains in the number of new patients when compared to the Quality designation. This is likely because the comparison of Premium to Quality physicians is a comparison of the highest ranking to the middle ranking, whereas the comparison between Premium and Not Designated physicians is a comparison of the highest ranking to the lowest ranking, so potential effects in the latter comparison are larger.

The design is well-suited to identifying heterogeneity of the effect of the designation on consumer demand because physician characteristics are continuous across the quality and cost efficiency thresholds. Premium designated physicians are comparable to Not Designated physicians on everything except designation when their underlying quality score is close to the quality threshold. Any difference in the characteristics of the new patients these physicians see represents heterogeneity based on patient demand responses to the disclosure program rather than physician characteristics.

I find that younger patients respond more to physician designation than older patients: the average age of patients seen by Premium physicians is significantly lower (by 1.4 years) than the age of patients seen by Not Designated physicians. This difference in the age of patients who see a Premium or Quality Physician versus to those who do not can be interpreted as being due to heterogeneity in patient response. This is true so long as physicians on either side of the cutoff also behave similarly after scores are released. There is no empirical evidence that physicians change their behavior or game their score to achieve a Quality or Premium Care designation, perhaps because scores are calculated using multiple years of claims (so averages are difficult to move), or because it is difficult for a physician to know exactly how a small change in their behavior maps to a slightly higher score.

The regression discontinuity estimates show how patients respond to the information program and reveal that younger patients respond more, but they do not shed light on which patients benefit the most or generate the most savings from seeing higher quality, more cost-efficient physicians. To study the heterogeneous effects of physician characteristics on patient outcomes, I complement the RD estimates with estimates from a two-way fixed effects "switchers" design, analyzing impacts on total spending and preventable emergency room visits.

The two-way fixed effects strategy requires that the decision to switch physicians is exogenous to patient outcomes. I focus on individuals who change physicians the same year their previous physician either moved or exited the dataset so that the decision to switch is plausibly exogenous. As a physician's cost-efficiency score increases, I find that their patients' spending decreases, on average, while an increase in a physician's quality score has the opposite effect (patient's spending slightly increases, on average). The mechanisms for these increases and decreases are different. More cost-efficient physicians prescribe less expensive (but the same number of) services as less cost-efficient physicians. Conversely, higher-quality physicians prescribe more services when compared to lower-quality physicians. While cost efficiency and quality impact various spending outcomes, there is no significant evidence that physicians who score higher on quality or cost efficiency shift outcomes on emergency room utilization. Preventable emergencies do not significantly increase or decrease with a switch to a higher quality or more cost-efficient physician.

Next, I next turn to heterogeneity in patient's age, subsetting by age group. I find that younger patients (aged 18-35) experience no significant gains or losses in spending due to their physician's quality or cost-efficiency score, while middle-aged patients (aged 36-55) experience larger impacts of both cost efficiency and quality. When applying the estimates from the model to an average middle-aged patient that switches from a Not Designated to Premium physician, they are expected to spend $30 fewer dollars in annual spending (about 4% of average annual spending) than they would have otherwise. When the average younger patient makes the same switch, they experience a small and not statistically significant gain in annual spending of $0.28 (about 0.4 % of annual spending).

Younger patients are more responsive to the Premium designation status, but older middle-aged patients experience the highest gains in savings from switching to Premium physicians. These results suggest that improved information targeting or larger financial incentives for middle-aged patients to respond to the Premium designation status may allow for larger net savings without negative impacts on health, as measured by preventable emergency room visits.

This study contributes to the literature on quality disclosure (Dranove and Jin 2010) by showing that privately insured patients respond to information about primary care physicians. This literature focuses primarily on how patients respond to ratings about hospitals and facilities (Bundorf, et al. 2009), health plans (Wedig and Tai-Seale 2002), and cardiac surgery ratings (Yoon

2020), or on the implications of providers adjusting their behavior in the presence of these programs (Dranove et al. 2003, Vatter 2021). I instead focus on the effects of insurer-provided quality and cost-efficiency scores on demand for primary care physicians. It is important to understand responses to information about primary care physicians because they serve as gatekeepers to the healthcare system by making referrals and provide essential preventative care services. The most closely related research by Chartock (2021) shows that patient satisfaction information impacts consumer choice for similar physicians. My study examines the effects of objective measures of quality, mainly measuring adherence to clinical practice guidelines, that patients may respond to differently than measures of patient satisfaction.

This study also contributes to the quality disclosure literature's exploration of heterogeneity in disclosure effects in two ways. First, I consider several possible explanations for the age heterogeneity. This extends the research that quantifies similar patterns in heterogeneity: Beaulieu (2002) finds that younger consumers are more sensitive to price and less sensitive to physician networks than older patients, and Chartock (2021) finds that younger and healthier patients respond more to patient satisfaction scores. Second, my research provides empirical evidence that motivates learning about heterogeneity. The fact that middle-aged patients are more impacted by the quality and cost efficiency of their physicians gives researchers a reason to learn about heterogeneity in responses to information.

Finally, my study contributes to the literature which breaks down patient outcomes into patient- and physician-driven components. My results add more evidence to the literature, which finds that physician characteristics can impact patient outcomes[5]. Whereas researchers have generally focused on Medicare patients, (Sabety 2021, Fadlon and Van Parys 2020, and Kwok 2019) this current study focuses on the younger, privately insured population. My results showing that middle-aged patients are more strongly impacted than the youngest patients and indicate the importance of examining this population. A study focusing only on elderly Medicare patients may lead to different conclusions since results vary by age.

[5] There is a large literature focusing on heterogeneity in provider practice styles and the impacts of provider-specific practice styles on patient outcomes. Chan et al. (2022), Currie and MacLeod (2017), Fletcher, Horwitz, and Bradley (2014), Molitor (2018), Schnell (2017), Schnell and Currie (2018), and Van Parys (2016) establish that providers vary and explore explanations for heterogeneous practice styles such as training, altruism, skill level, or environmental factors. Currie and Zhang (2021), Dhalstrand (2021), Doyle et al. (2010), and Gowrisankaran, Joiner, and Léger (2022) use random or quasi-random assignments of physicians to patients to understand the impacts of physician practice styles on patient outcomes, while Grytten and Sorensen (2003) and Simeonova, Skipper, and Thingholm (2020), use fixed effects designs to separate patient-specific from physician-specific impacts.

The paper proceeds as follows. Section 2 provides background information on the insurer's quality and cost disclosure program. In Sections 3 and 4, I introduce the study's data and empirical design. Section 5 presents results regarding the success of the program in steering patients toward Premium physicians, and in Section 6, I explore the heterogeneity of the response by patient characteristics. Sections 7-9 discuss the data, empirical strategy, and results of the switcher analysis, which identifies the impacts of quality and cost efficiency on patient outcomes both on average and by age. Sections 10 and 11 conclude the study.

2. Context: Information Disclosure and the Premium Program

I study the decisions of privately insured patients who choose primary care physicians in the context of an information disclosure program run by their insurance company[6]. Each year, binary indicators of cost efficiency and quality constructed using medical insurance claims are assigned to physicians and disclosed publicly online, so that patients can use the information to choose a high-value physician: someone who provides high-quality, cost-efficient care. Each physician's cost efficiency and quality designations are shown on the insurer's online physician directory. When a patient searches for a physician in their insurance network using the online directory to find a physician, they will be presented with a list of physician profiles. Each profile contains the cost and quality information in the form of blue hearts: a physician with two hearts next to their name is one with both higher quality and higher cost efficiency. The words "Premium Care Physician" are displayed alongside the hearts for these physicians. A physician who has higher quality but is not cost-efficient only has one heart next to their name, and the words "Quality Care Physician" are displayed. Physicians who have not met the quality threshold are not differentiated based on cost: both have two grayed-out hearts next to their name along with the words "Does Not Meet Premium Quality Criteria." Figure 1 depicts the hearts and language displayed for each type of physician. The information disclosure program began in 2011, and each physician's status is updated in the summer of each year based on claims from the prior 2-3 years. I study the 2019 program year because 2019 is the earliest year of program data available.

[6] I use data provided by a large, national health insurance company, which operates in 46 states in the United States. The data provider approved data questions before research was conducted and reviewed the draft of the paper, but did not actively participate in the research process.

The insurance company uses medical practice guidelines from the National Quality Forum and other sources[7] to calculate a physician's quality designation. Using claims information, the company assigns each physician three numbers which together comprise the score. First, they assign the number of "opportunities" the physician had to practice in accordance with the guidelines. The number of opportunities is generally equivalent to the number of unique patient-disease treatment combinations presented to the physician. For example, a patient with diabetes might count for four opportunities: the opportunity to perform a blood sugar control test, the opportunity to prescribe appropriate blood sugar control medication, the opportunity to check feet for abnormalities, and the opportunity to perform an eye exam. The physician's compliance with the opportunities presented by each patient is compared to the median compliance rate (at the national level) for each specific opportunity. A modified chi-square test is used to determine whether the physician's compliance is significantly lower than the national median compliance level. If it is, the physician is given the “Not Designated" (has not met the quality threshold) designation. More specifically, for each physician, χ is calculated as shown in Equation 1,

$$\chi = \begin{cases} -1 * \dfrac{(C-B)^2}{B} * \dfrac{T}{T-B} & if\ C < B \\ \dfrac{(C-B)^2}{B} * \dfrac{T}{T-B} & if\ C \geq B \end{cases} \tag{1}$$

where *C* refers to the number of compliant measures, *B* to the number of compliant measures the physician would have had if they had practiced at the median compliance level, and *T* is the total number of measures assigned to the physician. If $\chi \leq -5.4118$, then the physician is considered a Quality Care physician. For ease of interpretation, I refer to these physicians as “higher quality physicians", and those who missed the cutoff as “Not Designated physicians" or “physicians who did not meet quality criteria.”

There are a few important notes to keep in mind regarding the construction of the quality score. First, the score is calculated within physician specialty, patient condition, and in some cases, patient age or severity of disease.[8] This means that physicians who see sicker patients due to other reasons, such as social determinants of health, may have lower scores. Next, the score calculation is

[7] The majority of guidelines are from the National Quality Forum. Other sources include National Committee for Quality Assurance, Center for Medicare and Medicaid Services, and Pharmacy Quality Alliance, among others.
[8] For example, the prescription of antibiotics for young children with upper respiratory infections is treated as a separate measure from the measure that considers the prescription of antibiotics for elderly adults. Measures are generally split by patient characteristics in this way when evidence shows that different treatments may be more or less applicable to different types of patients.

increasing in the number of opportunities assigned to the physician.[9] Therefore, it is for the physician to have patients who are chronically ill, but who are good at managing their chronic condition and taking their prescribed medication. This is because some measures included in the quality calculation measure medication adherence. The number of opportunities and chronic illness management generally improve with age. Third, the quality designation is given to any physician who does not perform statistically worse than their benchmark. Thus, even though the compliance criteria are based on the national median performance for each measure, 90% of physicians met the quality criteria in 2019. Likewise, cost criteria are based on the 75th percentile of performance, and about 30% are considered cost-efficient. Finally, the score is based on adherence to clinical practice guidelines. If the very best physicians treat their patients in ways that go against guidelines due to complications or patient characteristics that are not observable within claims, these physicians may appear to be lower-quality physicians. However, on average, a higher score generally means the physician is doing more things they are “supposed" to be doing in terms of following guidelines and providing evidence-based care.

The insurance company also uses information from claims to designate each physician's cost status. The insurer creates a measure of the total cost of a patient over multiple calendar years and attributes that cost to the physician with the most significant involvement in the patient's care. The total cost measure aggregates all allowed, qualifying payments made by the patient and the insurer to physicians that the patient saw. Qualifying payments are determined by specialty. For example, costs associated with the diagnosis of Allergic Rhinitis are included in the cost measure for Allergists. The total cost variable includes both the patient's and the insurer's portion of the payment. The total cost variable is then risk-adjusted using the patient's predicted risk score based on patient conditions and demographic characteristics. Like the quality calculation, a benchmark score is then calculated based on the 75th percentile cost level, and a Wilcoxon rank-sum test is used to determine whether the adjusted costs attributed to the physician[10] were significantly lower than the benchmark score. I will call physicians “cost-efficient” physicians if their cost score lies below the threshold, and “non-cost-efficient” physicians if their score lies above the threshold.

Each physician's online profile contains other information. The name of the physician is included, as well as the physician's main address and the distance from the patient to the physician's

[9] This was not a deliberate choice, but results from the chi-square test statistic calculation. Later iterations of the program introduced measures to adjust the metric for physicians with very low numbers of measures.
[10] The costs attributed to the physician are also weighted by expected cost so that non-cost-efficient patients and procedures carry more weight in the calculation.

office. If the physician has a profile on Healthgrades.com (a popular physician rating website, where ratings are based on patient satisfaction), the average star rating is shown. A patient can sort physicians based on gender, distance, or quality/cost status; however, the default sorting is based in part by designation[11].

3. Data – Physician Regression Discontinuity

I use data from the information disclosure program in 2019 to measure each physician's underlying quality and cost score as well as their binary, publicly available quality and cost designation. The insurer uses claims information to construct continuous quality and cost scores for each physician. If each score passes the given threshold, the physician is assigned either the "Premium Care" designation if they are a higher quality, cost-efficient physician, the "Quality Care" designation, if they are a higher-quality, non-cost-efficient physician, or "Not Designated" if they are a lower quality physician (regardless of cost). I use data on the underlying scores and on the resulting status of each physician. Table 1 displays the number and proportion of physicians with each status in the bottom two rows. Surprisingly, 90% of primary care physicians are above the quality cutoff: 50% have a Quality Care designation (non-cost efficient), and 40% have a Premium Care designation (cost-efficient). The first two rows of the table show underlying cost and quality scores for physicians in each group. As expected, quality scores are higher for Quality and Premium Care physicians than for lower quality physicians, and cost scores are lower for Premium Care physicians.

When exploring patient steering, the main outcome variable is the number of new patients seen by each physician each month and each quarter. To create this variable, I begin with a dataset for each month of claims with patient and physician identifiers and procedure (CPT) codes. If the procedure code used corresponds to a new patient visit[12], then the patient is labeled as a new patient. The new patient indicator only counts first visits: if the new patient returns to the physician in the same month, only one new patient visit is accounted for. For most specifications, I create a cross-sectional dataset by aggregating to the physician level, including new patient visits for the 3 months after the new quality scores were released in 2019[13]. For specifications that rely on variation over

---

[11] The order of search results is based on physician tier (member-specific physician ranking based on plan characteristics, physician designation and ACO affiliation), physician designation, scheduling, cost efficiency ranking (further broken down but not into fully continuous cost groups), Healthgrades rating, distance, and physician name (alphabetical).
[12] The list of CPT codes used to designate new patients can be found in Appendix Table 1
[13] Program data prior to 2019 are unavailable, so I focus on the 2019 update to avoid disruptions caused by the COVID-19 pandemic

time, I use physician-month-level data. Table 1 shows the number of total and new patients by status in the cross-section. Premium Care physicians tend to see slightly more new and returning patients.

I also create variables for the number of new patients with various other characteristics, such as the number of new, younger (age 18-39) patients, or the number of new, chronically ill patients. While some characteristics, such as age and gender, are available directly in the medical claims, the chronic illness indicator must be constructed. I use the methodology from Gruber and McKnight (2016) to construct chronic illness indicators. For each patient month, I determine whether, in the prior year, the patient was seen in an office setting for a diagnosis of one of six common chronic illnesses.[14] If they were, then the patient is designated as chronically ill for that month.

To explore mechanisms of heterogeneous responses to being designated as a Premium or Quality Care physician, I create measures of situations that may impact how patients search: whether the patient or physician recently moved, whether the patient is new to the insurance company, and how much they paid out of pocket for the visit. First, I denote whether a patient or their physician has recently moved based on whether the first two digits of their zip code or their physician's zip code matches the first two digits of the zip code they had during their most recent prior visit to a physician of the same specialty. Next, I create an indicator for whether the patient is new to the insurance company based on their enrollment dates. If the patient was not insured during 2018 but was insured during 2019, the year studied, the patient is considered "new to the insurer." Finally, I explore the amount each patient paid out of pocket for the visit. While a patient may not know the exact amount they will be billed for a given visit ahead of time, they likely have a general idea of what their visit will cost based on whether they have already met their deductible or out-of-pocket maximum, and whether the type of care is fully covered (for example, an annual wellness check is fully covered). Patients may search for cost-efficient physicians based on their expectations of how expensive their care will be.

Finally, I merge in two additional physician covariates: average ratings from Healthgrades.com, and physician medical school graduation year. The Healthgrades score is merged based on name and location of the physician. About 12% of physicians within the bandwidth successfully merged to Healthgrades data. While the match rate is quite low, matching exactly on name and location ensures that matches are more likely to be accurate. Graduation year can be

[14] I follow Gruber and McKnight (2016) in identifying the following chronic illnesses: diabetes, asthma, arthritis, affective disorders, and gastritis

merged directly on physician identification numbers (NPIs), thus, the match rate is higher at 41%. However, graduation year is only available for physicians who also treat Medicare patients.

4. Empirical Strategy: Identifying the Steering Effect

I use a regression discontinuity design to determine the impact of quality designation on the number of patients the physician saw over the 3 months after designations were updated. Physicians are split into groups based on whether their quality and cost scores exceed quality and cost cutoffs. The insurer assigns labels to physicians, giving higher-quality, non-cost-efficient physicians a "Quality Care" designation, and higher-quality, cost-efficient physicians a "Premium Care" designation. Lower-quality physicians are "Not Designated." These three statuses lend themselves to three comparisons: the comparison of Premium to Not Designated, the comparison of Quality to Not Designated, and the comparison of Premium to Quality.

I analyze these three comparisons by subsetting the sample of physicians in different ways. First, to determine the impact of Premium against Not Designated, I subset to cost-efficient physicians, who are eligible for Premium status because they have lower cost scores. Comparing cost-efficient physicians just to the left versus just to the right of the quality cutoff identifies the impact of being designated as a Premium Care physician relative to Not Designated. Likewise, comparing non-cost-efficient physicians just to the left versus just to the right of the quality cutoff identifies the impact of being classified with a Quality Care designation relative to Not Designated, since non-cost-efficient physicians achieve the Quality Care designation rather than the Premium Care designation upon passing the quality threshold. Finally, subsetting to high-quality physicians, I compare physicians just to the left versus those just to the right of the cost threshold. Those to the left are high-quality, cost-efficient Premium physicians (since a lower cost score is better in this case), while those to the right are higher quality, non-cost-efficient Quality Care physicians.

Each designation is displayed in the form of hearts. Thus, the first natural experiment which compares Premium to Not Designated physicians is also a comparison of physicians with two hearts to those with no hearts. The second natural experiment, comparing Quality to Not Designated physicians, compares physicians with one heart to those with no hearts. The third natural experiment compares physicians with two hearts to those with only one heart.

To measure the difference in outcomes between physicians whose quality or cost scores were just-above versus just-below the threshold, I estimate a local linear regression within a small bandwidth around the cutoff, within the relevant subset (e.g., only cost-efficient physicians). Specifically, I regress

$$New\ Patients_j = \alpha + \delta_1 X_j + \beta 1\{X_j > 0\} + \delta_2 X_j 1\{X_j > 0\} + \epsilon_j, X \in [-10,5] \quad (2)$$

where $New\ Patients_j$ measures the number of new patients seen by physician *j* within the first 3 months of designation publication. X_j measures the continuous, underlying quality score, which is normalized so that the cutoff is zero. β measures the impact of passing the quality threshold and being classified as a higher-quality physician. The process is repeated with X_j as the normalized cost score to estimate the impact of being classified as a cost-efficient physician (comparing Premium to Quality physicians).

To visually assess whether discontinuities exist, I plot outcomes over bins of the running variable: either cost or quality score. I partition the running variable into 20 quantiles on either side of the discontinuity within the bandwidth and display the average outcome within each bin. I also plot the regression discontinuity predicted values (lines on either side of the cutoff) to help visualize the exact size of the discontinuity. I use a bandwidth which extends to 10 points below zero on the left of the cutoff and 5 points above zero to the right of the cutoff for all quality regression discontinuity plots, and a bandwidth extending to 3 points on either side of the cutoff for plots where the cost score is the running variable. These bandwidths were chosen based on mean squared error optimal bandwidths (Calonico, Cattaneo and Titiunik 2014), which vary between specifications. Using fixed bandwidths of [-10,5] and [-3,3] allows for the bandwidths to stay the same across all specifications. Because data are sparser on the left than the right of the quality threshold (recall that only 10% of physicians are below the quality cutoff), the mean squared error optimal bandwidths for the analyses around the quality thresholds are calculated allowing for different bandwidth sizes on the left and right.

4.1 Validity Checks and Evidence Against Gaming

Interpreting β as a causal effect requires that physicians are not able to engage in "gaming" by selecting which side of the cutoff they are on. It is unlikely that physicians can game in this way. Physicians are given detailed information of their performance on various measures; however, they are not told how performance on each measure maps to their quality score. To back out the mapping, a physician would need to know the national median compliance rate for each measure, the exact number of patients they treated with each condition, and the algebraic formula to calculate the resulting quality and cost scores. While this information is all available to physicians in principle, it is unlikely they would spend the time to calculate exactly how much additional care to give to specific patients. Next, since scores are aggregated over multiple years of claims, it is difficult to move the overall score by changing behavior in the short term. Finally, for a physician to respond to the program by changing their behavior, the physician would have to run another diagnostic test or make another pharmaceutical prescription for a patient for whom they would previously have not done this test or prescribed medication. It is difficult to imagine a physician who knew he or she was overlooking these tests for specific patients but did not decide to correct this behavior until after they were notified of their low-quality status. Thus, if the physician had known of their noncompliance all along, why would they not have corrected this behavior earlier?

The best evidence against physician gaming, however, is empirical. Figure 2 depicts histograms for all three natural experiments (comparing zero to two hearts, zero to one heart, and one to two hearts) in the spirit of McCrary (2008). If physicians can manipulate their scores to improve their likelihood of being classified as Quality or Premium status physicians, we would expect to see missing mass just to the left of the cutoff, and additional mass just to the right of the cutoff. No such patterns are detected in the data. I also run the density tests outlined by Calonico et al., (2014), and find no statistically significant discontinuities in the density of physicians around the cost or quality thresholds. The figures from these density tests are displayed in Appendix Figure 1.

If physicians are as good as randomly assigned around the cutoff, we would also expect that physician characteristics are continuous at the cutoff. To assess continuity around the quality cutoff, I predict the number of new patients seen by each physician as a function of physician gender and ZIP code fixed effects on the cross section of data from the 3 months prior to the 2019 program update. I then estimate Equation 2 with the predicted number of new patients as the outcome, shown in Figure 3, for each natural experiment. I find no significant impact or visual evidence of a discontinuity. Appendix Table 2 shows the regressions used to predict new patients.[15]

I also estimate the regression in Equation 2 with each physician characteristic as the outcome, and report results in Table 2. In Table 2, I also measure impacts on graduation year and

[15] Graduation year is only available for physicians who see Medicare patients, and Average Healthgrades Score is only available for physicians who have a Healthgrades.com profile and were successfully matched to the Claims dataset on name and location. These variables are omitted from the prediction of new patients because they are only populated for a subset of variables; however, the number of predicted new patients using all five variables is also smooth around the cutoff.

average patient review scores from Healthgrades.com, which are only available for a subset of physicians. I find no evidence of imbalance on these characteristics.

4.2 Identifying Patient Versus Physician Responses Using Differential Timing

I use a difference-in-discontinuity approach to identify whether physicians or patients are responsible for any differences in outcomes seen. While intuitively patients would be more likely to respond to the information, in principle it is possible that physicians could respond as well, perhaps by turning away new patients or advertising more heavily. The approach leverages disclosure timing: physicians were notified of their status 2 months before updated statuses were released publicly online. If effects materialize before patients could have seen updated statuses but after physicians were notified, then results can be attributed to physician behavior, whereas if results do not materialize until after physicians were notified of their score, the impacts were likely due to patient responses.

I estimate the following regression equation, adapted from Grembi et al. (2016) to break down effects year-by-year in an event study style framework:

$$New\ Patients_{jt} = \sum_{k=-5}^{8} 1\{t = k\}[\beta_k 1\{X_j > 0\} + \gamma_{1k} X_j + \gamma_{2k} X_j 1\{X_j > 0\}] + \tau_t + \alpha_j + \nu_{jt} \quad (3)$$

where X_j represents the running variable, underlying quality or cost score, and τ_t and α_j are month and physician fixed effects. β_k estimates the differential impact of having passed the threshold in year *k* relative to the effect in the month before the scores were given to physicians. If β_ks are close to zero and not significant during the time when physicians knew their updated status but patients did not, then we can infer that physicians did not respond immediately to the updated information. If effects materialize only after scores were available publicly to patients, then one can infer that patients did respond and physicians did not under two assumptions. First, one must assume that impacts on physicians were constant over time so that the new information did not take more than 2 months to change the behavior of physicians. Second, one must assume that there are no complementarities between physicians and patients both knowing information that could impact physician behavior (physicians cannot have waited for the public disclosure date to change their behavior). See Grembi et al. (2016) for further discussion of these assumptions.

4.3 Patient Heterogeneity

Measuring heterogeneous responses to a physician's designation is not straightforward because the treatment is at the physician level. To determine whether certain groups of patients respond more than others, I use two approaches. First, I use patient-physician level data from the 3 months after the 2019 program update and explore differences in characteristics of patients seen by various groups of physicians. Since physician characteristics are held constant by the identifying assumption of the regression discontinuity design, any discontinuity in patient characteristics can be attributed to either patients responding differently based on their characteristics, or by physicians purposefully targeting specific types of patients. I estimate Equation 2 with the average age and chronic illness status of new patients as outcomes, regressing at the patient level to avoid any aggregation bias that might result from first collapsing to the physician level and then running the regression. I cluster standard errors at the physician level to account for any within-physician correlation in chronic illness status and age.[16] I additionally estimate Equation 2 on the number of new patients who are younger than 40 years of age and on the number of new patients who are older than 40 years of age[17] and likewise for chronic illness status. To determine whether changes are due to physician or patient changes, I also estimate Equation 3, exploring dynamic effects, on patient characteristics. These complimentary analyses provide another way of examining heterogeneity in responses to the physician's designation.

5. Steering Effect Results

5.1 The Steering Effect

I begin by assessing the first stage. I break physicians into cost efficiency groups based on whether the cost score exceeds the cost threshold. For cost-efficient physicians, passing the quality threshold constitutes a sharp RD: everyone whose quality score exceeds the cutoff has Premium status. For physicians who are not cost efficient, passing the quality threshold should lead to Quality status. However, this discontinuity is fuzzy because some non-cost-efficient physicians achieve Premium status instead of Quality status upon passing the quality threshold. This is because some higher-quality physicians who do not meet the cost efficiency criteria can still achieve Premium

---

[16] Running regression 2 "at the patient level" means estimating the following regression: $Y_{ij} = \alpha + \delta_1 X_j + \beta_1 1\{X_j > 0\} + \delta_2 X_j 1\{X_j > 0\} + \epsilon_{ij}, X \in [-10, 5]$, where Y_{ij} is the patient-level characteristic such as age or chronic illness status

[17] Privately insured adults are usually less than 65 years of age, so 40 years of age is roughly between 18-65.

status if a large enough proportion of their practice group achieves cost-efficiency criteria. This is true only for cost efficiency; a physician who does not meet the quality criteria is never assigned Quality or Premium status based on the performance of their group. Appendix Figure 2 displays these first-stage plots in Panels A and B.

I next turn to the main results. To determine whether the effect of passing the quality threshold is driven by Quality or Premium physicians, I break physicians down into two categories based on their cost-efficiency status. Plotting new patients over physician quality in Figure 4 reveals a discontinuity around the quality threshold for primary care physicians,[19] but only for the cost-efficient physicians who achieve Premium status upon passing the quality threshold. A line is plotted on either side of the threshold, within the bandwidth. Passing the quality threshold leads to seeing more new patients for cost-efficient, primary care physicians, who have two hearts on their profile. For non-cost-efficient physicians, who only have one heart on their profile, impacts are much smaller and not statistically significant.

Panel C of Figure 4 displays results for the regression discontinuity estimation around the cost threshold, comparing physicians with one heart to those with two hearts. There is no visual discontinuity for either specialists or primary care physicians, and the regression equation picks up a null result (all regression results are displayed in Table 5). The null result could come from the fact that patients view one-heart (Quality Status) and two-heart (Premium Status) physicians as being more similar to each other in terms of number of hearts than no-heart (low-quality) versus two-heart (Premium Status) physicians. Or, the result could be because the cost RD is fuzzy, so the comparison across the cost threshold includes some physicians on both sides of the threshold with a Premium status, which would attenuate RD results. One way to determine whether the fuzzy nature of the RD is attenuating results is to run an Instrumental Variables (IV) regression, where passing the threshold is used as an instrument for cost-efficiency status. The regression results show that having cost-efficiency status leads to an insignificant increase in the number of new patients seen by physicians who barely passed the cost threshold. The results of this regression are displayed in Appendix Table 3. The IV results are not statistically significant, so it is unlikely that the fuzziness of the discontinuity is hiding a significant impact. However, it is possible that the effects of single-heart increases are too small to be detected in the sample, whereas the large effect comparing two hearts versus zero is detectable by the sample size. Further discussion of the differences between achieving

[19] Impacts for specialists are displayed in Appendix Figure 2.

a single heart (Quality Designation) and achieving double hearts (Premium Designation) is in Appendix A.1

The fact that effects are driven by cost-efficient physicians makes sense for multiple reasons. First, since sorting is in part based on status, achieving Premium Status may increase the page rank of physicians substantially, possibly to the first page of search results. While Quality Status physicians also experience an increase in page rank, the increase is likely smaller, and the likelihood that the physicians' profile is shown on the first page is also smaller. Further, Premium Status physicians have two hearts displayed on their profiles, while Quality Status physicians only have one heart displayed. This means that passing the quality threshold for cost-efficient physicians increases the hearts displayed by two hearts, whereas passing the quality threshold for a non-cost-efficient physician only increases the hearts displayed by one heart. The effects shown can be considered as the effect of being classified with a given designation, which includes both the ranking effect and the information effect. Disentangling these two effects is beyond the scope of this study because data on page rank are not available.

Exploring the impacts on primary care physicians is important because most research on physician quality disclosure has exclusively focused on cardiac surgeons, as in Yoon (2020), Dranove et al. (2003), and Dranove and Sfekas (2008). If patients respond differently to information regarding primary care physicians, then research on specialists such as cardiac surgeons may not generalize. Since primary care physicians often serve as gatekeepers to specialists, it is important to understand what factors lead patients to choose their primary care physicians.

5.2 Patient Versus Physician Responses

In evaluating the cross-sectional results, one may wonder whether the increases in new patients are truly due to patient demand, as opposed to physician behavior such as turning away new patients or changing other policies that may impact the number of new patients a physician sees. The timing of the information policy can be used to determine whether patients or physicians are responding, since physicians were informed of their status about 2 months before the scores were publicly visible to patients. Figure 5 displays the results of the difference-in-discontinuity regression shown in Equation 3. Each β_k is plotted over time, and the two vertical purple lines show the timing of physician and patient disclosures. Results for cost-efficient primary care physicians are shown, since the impacts on more expensive physicians are limited, thus there is little concern of physician behavior impacting results for those specifications.

There is no effect (relative to the month before physician disclosure) of being classified as a Premium Care physician in the 2019 classification between February and June of 2019, before physicians or patients were notified of their designation. This also serves as a validity test, showing first that physicians who would just pass the quality threshold in the 2019 update were trending similarly in terms of new patients as physicians who would just fail to pass the quality threshold. If physicians were gaming the cutoff, we might instead expect to see some form of pre-trends in this figure. The fact that no pre-trends are observed therefore gives credence to the assumption that physicians near the cutoff are comparable.

During the 2 months when physicians knew their status but patients did not, the effects stay around zero, whereas after new physician statuses were displayed at the end of September, the effects began to materialize. By October of 2019, the full effect is realized. The dynamic effects show that patient demand is most likely the driver of the empirical results, since no effects are seen until after patients were able to access updated physician status. On the one hand, one cannot rule out that physicians changed something about their capacity or likelihood of taking on new patients in a way that results in lagged effects. However, since the average wait time for a patient appointment for primary care physicians is 40 days (Penn, et al. 2019), it is likely that if physicians influenced the number of new patients they saw, effects would have been visible at least by August.

Collectively, the results in this section show that patients respond to the information program, and that effects are concentrated in cost-efficient physicians, who have two hearts displayed on their profile (rather than one out of two) and have higher page ranks. The results are likely due to patient behavior since the effects do not materialize until after patients have access to the information.

6. Heterogeneity

To determine whether specific types of patients respond more than others, I estimate Equation 2 at the patient level on patient characteristics as outcomes. If certain types of patients respond strongly to a physician's designation, the average characteristics of patients treated by those physicians will change. Table 4 displays the results of regressions on average age, chronic illness status, and previous year's spending for new patients treated by primary care physicians who barely passed the quality threshold versus those who barely missed it. My findings show that passing the quality threshold leads to a relative increase in the average age of patients, and a marginally significant relative increase in healthy (non-chronically ill) patients. There is no impact on how expensive the patients seen tend to be (as measured by last year's spending).

The average age and chronic illness status of patients could be decreasing for primary care physicians either because young patients respond more to premium status, or because older patients respond negatively to premium status. To evaluate these possibilities, I display regression discontinuity plots in Figure 6 separately for new patients who are young (18-39) versus new patients who are older (40 and above). I standardize the measure of new patients for each group since the base rate of new patients is different between groups, so changes in the standardized measure can be interpreted as proportional increases from the mean of each outcome. Figure 6 depicts that while both older and younger patients respond positively to the status of cost-efficient primary care physicians, younger patients respond more. Likewise, both chronically ill and healthy patients respond, but healthier patients respond slightly more than chronically ill patients. Figures displaying the regression discontinuity plots for non-cost-efficient physicians are in Appendix Figure 5.

To interpret these impacts as being due to patient demand rather than changes in physician behavior, I estimate Equation 3 at the patient level on patient outcomes. If patient characteristics change before patients can view updated scores, then the differential impacts may be due to physicians changing the types of patients they see rather than different patients responding differently to the information. Appendix Figure 4 displays impacts on average age and chronic illness for cost-efficient primary care physicians over time. My findings show that impacts on average age are significantly lower during post-disclosure months. Impacts on average chronic illness status are less conclusive: there is some evidence of even slight relative increases in the level of chronic illness; however, some pre-periods have significant increases as well. For age, responses are likely driven by patient behavior, but heterogeneity in chronic illness status does not hold up to this robustness test. Therefore, I focus only on how a patient's age impacts their response to the information program in the remainder of this paper.

6.1 Explaining Heterogeneity

Younger patients respond more to premium status than older patients do when they see primary care physicians. Is this due to the search behaviors specific to younger patients versus older patients, or do other characteristics that correlate with age and health impact search behavior? Understanding drivers of age-related heterogeneity is important so that policymakers can more effectively target information in future programs.

First, I evaluate one possible reason for the heterogeneity: that older patients are more attached to their physicians than younger patients are, which leads to reduced switching for older patients. Older patients may be more attached to their physicians, leading fewer of them to switch. This is true in the dataset: the average number of older new patients is smaller than the average age of younger new patients. However, if this difference explained the heterogeneity (i.e., that responses were proportionally the same, but the base rates are lower for older patients), then the standardized number of new patients displayed in Figure 6, which nets out the average new patients for each group, would not display differences in effects for older versus younger new patients. Further, the impacts on average age displayed in Table 4 already adjust for different base rates since those base rates play a role in the average age of patients on both sides of the cutoff. If different base rates did explain results, then no impacts would be found. Since different base rates do not explain results, the heterogeneity must be explained by something else.

Next, I evaluate whether three search-related characteristics explain the age and health-status related heterogeneity. While younger patients are more likely to have recently moved and more likely to be newly insured(as shown in Figure 7), there is no substantial heterogeneity based on these characteristics which one might otherwise have thought would impact the search process. Individuals who recently moved may have a smaller network to rely on when searching for a physician and may therefore have to rely more on website information. Individuals who are new to the insurer may have less exposure to the program and therefore may not know about the program or know how to search through the insurer's online website. Table 5 displays results of estimating Equation 3 at the patient level on outcomes for whether the patient or their physician recently moved, whether they were new to the insurer in 2019, and the amount spent by the patient for cost-efficient primary care physicians. My findings show no significant increases or decreases in the proportion of patients who recently moved or who were new to the insurer; however, there is a significant increase in the amount paid by patients who saw cost-efficient physicians who just exceeded the quality score cutoff.

The increase in the amount paid by patients who saw Premium (higher quality, cost-efficient) providers who just passed the cutoff suggests that financial considerations may explain heterogeneous effects by patient age. This result should not be interpreted as information about what happens to patient spending when they see a Premium physician because the Premium physicians who just passed the quality threshold are comparable in cost efficiency (and other underlying characteristics) to those who just missed the threshold. Instead, the impacts should be interpreted as identifying differences in the types of patients (those who were about to spend more versus less) who choose to see Premium versus Not Designated physicians. Since age is negatively correlated with out-of-pocket spending, it could be the case that young patients are more responsive to information about physician costliness than older patients because they have more to gain in terms of cost savings. This hypothesis is further supported by the fact that heterogeneity based on age is only present for cost-efficient physicians, whereas young and old patients respond similarly to the Quality Care designation (see Appendix Figure 5).

7. Data - Switchers

I now turn to the switcher analysis to understand the downstream outcomes of switching to higher quality, more cost-efficient physicians. This analysis requires a panel of patients who are each matched to a single physician. To create this panel, I use claims data from patients who were insured between 2015 and 2021. Since effects are limited to primary care physicians, I limit my analysis to specialties of Family Medicine and Internal Medicine who were subjected to the disclosure program.[20] Each patient is matched to their modal primary care physician: the physician they saw for the largest number of visits in each year. Approximately 73% of patients had two or more modal primary care physicians. I keep these in the analysis dataset, and drop any patients who did not visit any physician over the full 7 years included in the dataset[21].

To create a more exogenous measure of physician switching, I keep only patients who switched modal physicians the same year that their original physician either exited the dataset (retired or became out-of-network for the insurer) or moved far enough away that the first two digits of their ZIP code changed. These patients were likely forced to switch physicians, rather than making the decision to switch based on some characteristic related to their outcomes. The final dataset includes 1,385,461 observations for 197,923 patients over 7 years[22].

[20] Program participation is largely determined based on specialty and number of patients seen. If an internal medicine physician primarily practices in a subspecialty which is not covered by the program (say, palliative medicine or sports medicine) they will be omitted. Physicians who do have appropriate subspecialties but see fewer than 20 patients will also not be included.

[21] I examine robustness of results using a panel that is balanced on the year relative to the switch rather than the calendar year.

[22] This dataset includes patients who interacted with the health system every year (though continuous enrollment is not specifically required), whose modal primary care physician is eligible for the Premium program, who only have one primary care physician and see an eligible primary care physician each year. The total number of patients for each of these subsets is displayed in Appendix Table 4.

This analysis explores the impacts of quality and cost efficiency on both quality and cost-efficiency-related measures of healthcare utilization. I first study measures of spending, exploring the impact of physician quality and cost efficiency on the total amount paid by both the patient and the insurer for care provided by the modal primary care physician. I also explore the total amount paid over all physicians the patient saw in a year to determine whether there are any spillovers in cost efficiency. To determine whether patients are better or worse off after having switched physicians, I also examine the total out-of-pocket payments made by the patient for their modal primary care physician and over all physicians.

To understand whether impacts on spending based on quality and cost efficiency are driven by volume of services or the price of services, I explore impacts on four outcomes: the number of services the patient received that year (measured by the number of unique service lines on the medical claim), the number of visits the patient had with their modal physician, the average number of services per visit (dividing the number of services by the number of visits) and the average "price per visit," the amount paid by the insurer and the patient divided by the number of visits.

Next, I examine emergency department utilization. I follow Alexander et al. (2019) in separating emergency room visits into three categories: visits for true emergencies that are preventable, to some extent, by appropriate primary care ("preventable emergencies"), visits for non-emergency care ("unnecessary visits"), and visits for true emergencies that are not primary-care preventable (called "placebo visits", since these should not change with better or more cost-efficient primary care).

Preventable emergencies are measured as any emergency room visit[23] for asthma, diabetes, influenza, heart attack, angina, or stroke. Asthma attacks, diabetes complications, and heart attacks can all be prevented to an extent by appropriate prescriptions and training from the physician on how to manage the condition, and influenza emergency visits may be prevented by an annual flu shot. Of course, much of the management falls to the patient, but part of being a higher-quality physician is being able to teach patients how to appropriately manage their chronic conditions.

While preventable emergencies are generally regarded as negative outcomes, it is possible that higher-quality physicians could increase preventable emergencies by changing the threshold at which a patient decides an emergency visit is necessary. For example, a very high-quality physician

[23] Emergency room visits include any claim with either an emergency room place of service code (23), an emergency room revenue center code (450-458), or emergency room CPT codes (G0380-G0384). These capture emergency room visits regardless of whether the patient was subsequently admitted to the hospital.

might communicate clear guidelines about when an emergency room visit is appropriate, while a lower-quality physician may not. Under the care of a higher-quality physician, this could result in a patient going to the emergency room more frequently, if their health issues were always above the threshold for seeking emergency care, but the patient did not know that until after having seen the higher-quality physician.

I measure unnecessary visits as emergency room visits for primary-care treatable conditions: urinary tract infections, conjunctivitis, upper respiratory tract infections, sore throat, and ear infections. These are visits that would most appropriately be treated in a primary care or office setting, rather than an emergency department. Higher quality and more cost-efficient physicians should decrease these visits, redirecting care to the appropriate setting.

To measure placebo visits, I count the number of visits due to childbirth, poisonings, and fractures. These conditions need to be treated in an emergency room setting but are not outcomes a physician has control over. No increase or decrease is expected with a switch to a higher quality or more cost-efficient physician.

Table 6 shows summary statistics over all outcomes for physicians broken into three categories of patients: "Stayers," those who continued to see the same primary care physician over the full panel, "switchers," who switched physicians at least once, and "induced switchers" whose physician switch coincided with the year their former physician moved or exited the dataset. The first nine rows in Table 6 display the mean and standard deviation of spending and emergency room utilization variables, and the following rows display differences in physicians' average quality and cost-efficiency scores as well as the average change in these scores for patients who switch physicians. Across all outcomes and quality and cost efficiency variables, stayers, switchers, and induced switchers look similar.

8. Empirical Model – Switchers

I exploit patient switches between physicians to determine the impact of quality and cost efficiency on utilization outcomes. Adapting the model from Finkelstein et al. (2016) to my setting, I define the patient's utility as a function of their utilization conditional on health status h_i and a preference parameter η_i. There is an appropriate level of utilization given the patient's health status which is adjusted by the preference parameter. A patient with a higher level of η_i prefers more utilization; perhaps they use the emergency department more frequently or prefer physicians to prescribe more invasive or expensive procedures.

$$u(y|h,\eta) = -\frac{1}{2}(y - h_{it})^2 + \eta_i y \qquad (4)$$

The physician maximizes their own utility, $\tilde{u}$, which includes patient utility adjusted by the physician's own quality and cost efficiency parameters. Both quality, λ_j^q, and cost efficiency, λ_j^c, impact how much care physicians give to their patients, and may impact care in different ways. For example, a higher quality physician may run more tests, leading to higher utilization, whereas a more cost-efficient physician may choose less expensive tests. By assumption, quality and cost efficiency are separable.[24] The physician also faces time-varying costs of providing care, PC_{jt}.

$$\tilde{u} = -\frac{1}{2}(y - h_{it})^2 + \eta_i y + \gamma_j^c y + \lambda_j^q y - PC_{jt} \qquad (5)$$

The physician chooses the level of provision to maximize their utility function. The physician's optimal choice of utilization is a function of patient-specific parameters, physician cost efficiency and quality, and a time-varying component. I assume that PC_{jt}' is linear in utilization and additively separable in j and t, and that the physician component of PC_{jt}' is captured by λ_j^c. This assumption implies that the physician's cost efficiency parameter reflects the physician's own propensity to provide cost-efficient care and how costly it is for that physician to provide care. The separability assumption allows PC_{jt}' to be displayed as the linear combination of λ_j^c and a time fixed effect τ_t. I next assume that the level of utilization which maximizes $-\frac{1}{2}(y - h_{it})^2$ is comprised of a patient fixed effect, $\hat{\alpha}_i$, and a set of observable time-varying controls, x_{it}, which include relative switch-time indicators. These indicators capture variation in optimal utilization based on the amount of time until a switch, allowing for the decision to switch physicians to be correlated with changes in health status over time. Under these assumptions, first order conditions from a maximization of the utility in Equation (5) imply that utilization y_{ijt} for a patient i who saw physician j in year t is:

$$y_{ijt} = \hat{\alpha}_i + \tau_t + \lambda_j^c + \lambda_j^q + x_{it}\beta + \epsilon_{ijt} \ (6)$$

[24] The separability assumption can be relaxed by modeling the utility as $\tilde{u} = -\frac{1}{2}(y - h_{it})^2 + \eta_i y + \gamma_j^c y + \lambda_j^q(1 - \lambda_j^c)y - PC_{jt}$. The interaction of λ_j^q with $(1 - \lambda_j^c)$ can be interpreted as higher quality physicians leading to higher utilization (perhaps through increased testing), while higher quality physicians are also less biased away from the "optimal" level of utilization (that which maximizes $-\frac{1}{2}(y - h_{it})^2$) by their own cost efficiency. In other words, a low-quality, cost-efficient physician might skimp on all tests, whereas a high-quality, cost-efficient physician would run only the important tests. Results in Appendix Table 5 imply no such complementarities exist in the data, so the separability assumption is likely valid.

I then rewrite Equation (6) as a function of pre- versus post-switch timing in Equation (7) as follows:

$$y_{it} = \hat{\alpha}_i + \tau_t + \lambda_o^c + \lambda_o^q + 1\{Post\}(\theta^Q \Delta Q + \theta^C \Delta C) + x_{it}\beta + \varepsilon_{it} \quad (7)$$

where $\Delta_Q = Q_d - Q_o$ and $\Delta_C = C_d - C_o$, the difference in physician scores for quality and cost efficiency between the "destination physician" (d) that the patient switched to and the "origin physician" (o) that the patient switched from. Note that in Equation (7), the physician scores (Q, C) are not equivalent to the physicians' costliness and quality parameters λ_Q, λ_c. These scores instead include both physician-specific and patient-specific components of measurable costliness and quality. The parameters θ^Q and θ^C in Equation 7 therefore represent the ratio of the difference in quality and costliness to the difference in measurable costliness or quality scores: $\theta_c = \frac{\lambda_d^C - \lambda_o^C}{C_d - C_o}$. These parameters can then be interpreted as the increase in utilization which is attributable to physicians per a one-unit increase in costliness or quality scores. These need not lie between zero and one because the units of the numerator and denominator are different.

Finally, note that the characteristics of the origin physicians, λ_o^c and λ_o^q, are captured empirically by the patient fixed effect, so that the regression I run is displayed in Equation 8,

$$y_{it} = \alpha_i + \tau_t + 1\{Post\}(\theta^Q \Delta Q + \theta^C \Delta C) + x_{it}\beta + \varepsilon_{it} \quad (8)$$

where $\alpha_i = \hat{\alpha}_i + \lambda_o^c + \lambda_o^q$.

The model implicitly assumes that outcomes depend only on the current physician, and not the history of physicians that the patient has seen. For most patients, this assumption is satisfied because the utilization outcomes studied are quick to come to fruition. For example, corticosteroids used for the treatment of asthma take only 4-6 weeks to improve breathing.[25] If the assumption is not met, then θ^Q and θ^C simply measure deviations from the lasting impacts of the prior physicians, which are absorbed by the patient fixed effects. The model also implicitly assumes homogeneity in treatment effects: that θs do not vary with patient characteristics. This assumption will be relaxed in various specifications which explore impacts within older and younger patients.

Recent literature has shown that heterogeneity over treatment timing, where earlier-treated groups respond differently than later-treated groups, can also bias results. In Appendix Section A.2, I explore robustness to this concern using novel estimators which relax the assumption of no heterogeneity over treatment timing.

[25] https://www.mayoclinic.org/drugs-supplements/corticosteroid-inhalation-route/proper-use/drg-20070533

To interpret θs as causal parameters, the post-switch indicator as well as Δ_Q, and Δ_C must be exogenous. This means that there cannot be any unobservable time-varying characteristics that correlate with both the outcome and either the post-switch indicator or the change in quality or costliness of the physician. There is evidence in this study that different types of patients are more or less likely to choose high-quality, cost-efficient physicians; however, these patient types are generally fixed over time. Therefore, the correlation is captured by the patient fixed effect. To increase the plausibility of exogeneity, I focus on the subset of patients who switch physicians because their previous physician either moved or exited the dataset. I call these "induced switches." These induced switches were forced, ruling out patient selection in the decision to change physicians.

While the assumption of parallel pre-trends is not necessary or sufficient for the interpretation of thetas as causal parameters (Hull 2018), it is still helpful to visualize trends in spending for patients switching to higher versus lower quality and cost-efficient physicians. If trends were markedly different leading up to the switch year for those who switched to higher quality physicians, it would be more difficult to justify the assumption that there is no selection on time-varying unobservable characteristics into higher quality physicians. To visualize pre-trends, I regress the following, estimating separate parameters for each year relative to each patient's first switch.

$$y_{it} = \alpha_i + \tau_t + \sum_{k=-6}^{5} 1\{Year = k\}\left(\theta_k^Q \Delta Q + \theta_k^C \Delta C\right) + x_{it}\beta + \varepsilon_{it} \quad (9)$$

9. Impacts on Downstream Outcomes

In this section, I explore the results of the switcher regressions in equations 8 and 9, both on average and broken down by patient age.

9.1 Average Effects

I estimate Equations 8 and 9 for patients who saw primary care physicians between 2015 and 2021 on measures of utilization. First, I explore total spending, analyzing the amount paid by both the insurer and the patient each year. Figure 8 displays the results of estimating Equation 9 on total spending. Panel A reports the coefficients on a one-point increase in cost efficiency, while Panel B reports the coefficients on a one-point increase in quality. In both panels, pre-trends are relatively flat, with individuals who switch to higher cost-efficiency physicians trending similarly, and likewise for quality. After the patient switches physicians, spending experiences a relative decrease for every additional cost efficiency point, and a relative increase for every additional quality point. Figure 8 includes data on all patients who saw a primary care physician who was part of the disclosure program during each calendar year; however, this mechanically means that the panel is unbalanced on time relative to the switch. To overcome this issue, Appendix Figure 6 displays the impacts using a dataset that is instead balanced on relative time: the years leading up to and after a switch. Balancing on relative year removes numerous observations, so effects are underpowered; however, they are similar in magnitude, so it is unlikely that panel imbalance on relative year biases effects.

Table 7 shows the results of the regression in Equation 8 on a variety of spending outcomes. Column 1 reports the impacts of a single-point increase in quality and cost efficiency. A single-point increase in quality increases payments by $1.66, while a single-point increase in cost efficiency decreases payments by $11.18. Appendix Table 5 shows the results of these regressions including an interaction term between cost efficiency and quality, and there is no significant impact from the interaction, adding credence to the assumption that cost efficiency and quality are separable.

To put these impacts in perspective, when the average patient switches from "Not Designated" to "Premium" physicians, the quality score increases by 2.14, and the cost efficiency score increases by 1.80. Multiplying these by the impacts of single-point increases predicts that for the average switch from Not Designated to Premium physicians, a patient saves $16.56.

Table 7 also examines mechanisms by which spending may increase or decrease, particularly by breaking total spending down into quantity (number of services received) versus price (amount spent per service received). Columns 2 and 3 in Table 7 display impacts of switching to higher quality and more cost-efficient physicians on the number of services and services per visit, while Column 4 shows the impacts on the average price per service: the total amount paid divided by the number of services.

Interestingly, the mechanisms for impacts on spending are different for quality and cost efficiency. Spending increases that are associated with switches to higher quality physicians are due to increases in both the number of services in total and on a per-visit basis, with single-point increases in quality leading to small but significant increases in the number of services the patient received, but if anything, slightly less expensive services being done. On the other hand, more cost-efficient physicians seem to achieve cost efficiency by charging lower prices, where single-point increases in cost efficiency leads to $1.61 less being spent per service received. While these analyses break down quantity and prices, they cannot capture one facet of the change in average price per service: whether the decrease in price per service comes from the exact same services costing less (possibly through different results of negotiation over allowed prices), or whether it comes from more cost-efficient physicians prescribing less expensive services (i.e., choosing inexpensive medication instead of surgery).

Columns 5 and 6 of Table 7 display impacts on the total amount paid over all physicians an individual saw each year, not just their modal primary care physician in Column 5, and the amount paid for services received by physicians other than their modal PCP in Column 6. The results in these columns show that when a patient switches to a more cost-efficient physician, their overall spending across all physicians decreases. This is not just because the PCP's contribution to total spending is high, but because spending associated with non-modal physicians also decreases. This could be because more cost-efficient physicians refer to more cost-efficient specialists. However, it could also be that the cost-efficient physicians prescribe less expensive lab tests, but other physicians who interpret the results of the labs are noted on the claim rather than the prescribing physician.

The final column of Table 7 displays the impacts on patients' out-of-pocket spending, a subset of total spending. The results in this column show that switching to a more cost-efficient physician results in a marginally significant decline in out-of-pocket spending. The proportional impact is the same as the impact of cost efficiency on total spending: the $11 decrease on a base of $680 in total spending is almost identical proportionally to a $3 decrease on a base of $185. Thus, these results do not imply that the insurer receives all the gains to cost efficiency, but more likely that the magnitudes of the effects on patients' out-of-pocket spending are too small to be detected by the sample size.

I next explore outcomes on emergency room utilization. In Table 8, I report the results of estimating the regression in equation 8 on three measures of emergency room utilization: preventable emergencies, non-emergencies, and placebos. Column 1 shows the impacts on preventable emergency room visits, which are true emergencies that are in part preventable with high-quality primary care. Column 2 shows impacts on unnecessary emergency room visits, which are not emergencies and are more appropriately treated in an office setting, and Column 3 shows impacts on placebo emergencies, which are true emergencies that cannot be decreased or increased through improved primary care.

Table 8 shows no evidence of any statistically significant impacts of quality or cost efficiency on emergency room utilization, except for a marginally significant increase in the measure of placebo visits for more cost-efficient physicians. I examine the marginal increase in placebo visits further and find it is not robust (see Appendix section A.3). The 95% confidence intervals on the effect of a one-point increase in quality range from a decrease of 0.017 emergency room visits to an increase of 0.016, and the effect of a one-point increase in cost efficiency range from a decrease of 0.007 emergency room visits to an increase of 0.007. These confidence intervals are fairly precise, including a 1.2% increase and 2% decline in preventable emergencies for a one-point higher quality physician and an 8% increase and an 8% decrease in preventable emergencies for a one-point more cost-efficient physician.

One concern about using preventable emergencies as the main health outcome is that these rare outcomes may be difficult to move. However, other interventions have been shown to significantly impact these outcomes. Alexander et al (2019) find that retail clinics can significantly decrease both preventable and unnecessary emergency room visits, and Miller (2012) finds that having health insurance leads to fewer unnecessary emergency room visits. To improve the statistical power of results, I also explore impacts when including endogenous switches (see Appendix Table 5). My findings show that these impacts are similar in magnitude, but the effects are more precise, with confidence intervals on quality spanning a 0.3% decrease to a 0.2% increase in emergency room visits, and confidence intervals on cost efficiency spanning a 1.9% decrease and a 1.4% increase.

9.2 Heterogeneous Effects

I next separate patients by age group, calculating the impacts of changes in quality and cost efficiency on outcomes separately within each group. Table 9 displays the impacts for six 10-year age bins for patients aged 18 and older. Column 1 shows impacts for the youngest patients, aged 18-25, and age increases up to Column 6, which shows impacts for those who are 65 and older.

Columns 1 and 2 show no significant impacts of quality or cost efficiency on total spending for the younger patients, while Columns 3 and 4 show both significant increases in spending due to increased quality and significant declines in spending due to increases in cost efficiency, respectively. The impacts then decline again for patients who are 56-64, and while patients 65 and older do show large point estimates, none are statistically distinguishable from zero. Multiplying the impact of cost efficiency and quality by the increases in quality and cost efficiency scores for the average (within-age-group) increases in quality and cost efficiency from switching from Not Designated to Premium physicians, patients who are 46-55 years old stand to save nearly $30 in the year of the switch, whereas the youngest patients would see a not statistically significant spending increase of $0.28.

When the average younger patient makes the same switch, they experience a small and not statistically significant gain in annual spending of $0.28.

The number of patients in the dataset varies with age, with only 658 observations in the youngest age group, and up to approximately 16,000 in the older age groups. One may wonder if results for the younger patients are underpowered to detect significant changes in spending. One way to examine this hypothesis is to include all switches, including the switches that are not induced by physician moves or exits. The assumption that switches are exogenous is less plausible for this larger sample; however, there are many more switches to learn from. Appendix Table 9 displays the results from running the switcher regression on the full set of physician-patient switches. Impacts of cost efficiency for all age groups are larger, so the inclusion of endogenous switches may bias results away from zero. However, the pattern of effects is the same, with largest impacts for the middle-aged patients who are 36-55, and smaller impacts for other patients.

The general pattern of results remains when analysis is done on the subset of the data which is balanced on year relative to the switch, rather than calendar year. The subset contains patients who exist in the dataset from 2 years before the switch until 2 years post-switch. The requirement of having 2 years post-switch included in the data means that only switches from 2017-2019 can be included (i.e., 2020 and 2021 are the two post-switch years; 2015 and 2016 are pre-switch years). This decreases the size of the dataset considerably but removes any concern that longer-run impacts are driven by changes in the sample rather than true dynamic effects. Appendix Table 10 displays impacts over age groups on this subsample. The general pattern of results remains, with the 36-45 year old patients experiencing the largest benefits to switching to cost-efficient physicians. However, the impacts of physician quality are less robust.

In a final robustness check, the dynamic effects within age groups are shown Appendix Figure 7. The purpose of this check is to ensure that within age groups, patients who switch physicians earlier are otherwise trending similarly to patients who switch later. This check rules out the possibility that different effects are driven by different trends within age groups.

Nonlinearity over patient age may also be present for health outcomes, even though no effects on preventable emergencies were found on average. Table 10 shows the impacts by age group on preventable emergency room visits. There are no statistically significant impacts of quality on preventable emergency room visits, regardless of a patient's age. There are two age groups where statistically significant impacts on cost efficiency are detected. Patients 46-55 experience increases in preventable emergencies when they switch to more cost-efficient physicians, and patients 56-64 experience marginally significantly fewer preventable emergencies.

I again examine robustness of these results to both inclusion of endogenous switches in Appendix Table 11, and subsetting only to a relative-time balanced panel in Appendix Table 12. Appendix Table 11 shows that when all switches are analyzed, the marginally significant decline in preventable emergencies remains for patients aged 56-64. However, the significant increase for patients 46-55 disappears. Additionally, there is a marginally significant increase in preventable emergencies for the youngest patients. Appendix Table 12 shows that the decline in emergency room visits for ages 56-64 is again robust; however, for ages 46-55, the significant increase in preventable hospitalizations is no longer there, while a new significant decline in preventable emergencies shows up for the same age group (for a one-point increase in the quality score). Together, these exercises suggest that the decline in preventable emergency room visits for 56-64 year-old patients is robust, but the increase in preventable emergency room visits from switching to more cost-efficient physicians for 46-55 year-old patients is possibly driven by changes in dataset composition. Overall, the analysis points to decreases on spending being larger for middle-aged patients, and marginally significant decreases on preventable hospitalizations for the 56-64 age group.

9.3 Mechanisms

The effects are concentrated on middle-aged patients, and there are a few reasons why this may make sense. First, middle-aged patients are likely to be newly diagnosed with a chronic illness. Appendix Figure 8 shows that the likelihood of having a current diagnosis that did not exist in the previous year for any chronic illness has an inverted U-shape over age, peaking between ages 40 and 60. Table 11 explores whether the onset of chronic illness explains the fact that middle-aged patients are more heavily impacted by the quality and cost efficiency of their physicians, separately estimating impacts over the age distribution for patients who became newly chronically ill during the sample period. Column 1 of Table 11 displays impacts on the amount paid for younger patients, aged 18-35, who have no change in their health status, whereas Column 2 displays results for those who became chronically ill at some point during the sample period. Columns 3 and 4 display these results for middle-aged patients, while Columns 5 and 6 display the same for the oldest patients.

While the onset of chronic illness is correlated with age in the same way treatment effects are, it does not explain the pattern of results over patient age. If the onset of chronic illness did explain results, then one would expect to see significant effects across the age distribution for the newly chronically ill. This is not the case. There is some evidence that the impact of cost efficiency for middle-aged patients is driven by the newly chronically ill; however, the point estimate for healthy middle-aged patients is just as large, but noisier. There are also some differences noted in the impact of quality on spending which differ by health status, but these patterns do not imply that health status explains the pattern of results over patient age.

There is no evidence that the onset of chronic illness explains the patterns over patient age. However, there are many other characteristics that vary with age and could instead explain the results, many of which cannot be measured without additional data. For example, younger patients in this dataset may be more diligent about their health (they visit their physician's office annually despite being quite healthy), whereas older patients may be less diligent and therefore more reliant on their physician to make good treatment choices. Age could also vary with factors such as income or socioeconomic status (SES) to the extent that selection into private insurance is impacted by these factors differently by age. Diligence and SES could therefore play a role in explaining the differential impacts by age.

10. Discussion and Conclusion

This study aimed to answer two research questions, (1) determining which patients respond to disclosure programs, and (2) determining which patients benefit most from responding to disclosure programs. My findings show that younger patients respond more to the signal of higher quality and cost efficiency than do older patients, so the average age of new patients seen by Premium physicians is significantly lower than the average age of patients seen by Not Designated physicians. On the other hand, it is the middle-aged patients, not the youngest patients, who experience the largest cost savings from seeing higher quality, more cost-efficient physicians.

Figure 9 compares the estimates of the two sections of the paper, first plotting average steering effects by age, showing that patients who are 18-35 respond the most to disclosure programs. Premium physicians attract relatively more new patients from these age groups than from other age groups. In the lighter colored bars, I plot the impacts of switching to a more cost-efficient physician (for a one-point increase in score) separately by age group and find that and impacts are largest for the middle-aged patients who are 36-55 years old. Clearly, the patients who are most impacted by switching to more cost-efficient physicians are not the same patients who respond to the Premium designation status.

Comparing results from the regression discontinuity design to those from the two-way fixed effects switchers design requires a number of assumptions. First, to interpret the regression discontinuity effects as the effect of being classified as Premium regardless of underlying quality level, one must assume that the treatment effect does not depend on the physician's quality or other characteristics that vary with the quality score. One may wonder whether quality correlates with capacity such that higher quality physicians may experience lower gains to the Premium designation because they are already close to capacity. While the findings show that impacts do vary with physician capacity (see Appendix Figure 10), capacity does not vary substantially over the quality distribution. Second, one must assume that the young patients who are steered by the Premium designation toward Premium physicians are comparable to the young patients in the switcher analysis panel. The switcher panel includes patients who interacted with the health system every year, so these patients are possibly sicker than young patients or are more diligent about their health (always going for annual checkups). On the other hand, the steering dataset does not subset based on patients' repeated use of the health system. When these assumptions are met, one can compare impacts and find that young patients respond to the Premium designation, whereas older middle-aged patients experience larger savings.

Savings can be maximized when patients who have the most to gain are paired with the highest-scoring physicians. In this light, it would be wise to consider a policy that would re-sort patients so that the middle-aged patients (those who have the most to gain) are paired with the most cost-efficient physicians. I simulate such a policy below. While this exercise ignores several important details (such as what type of policy could lead to this sorting, and other inputs to total welfare such as physician-to-patient distance, and the patient's preferences over other physician characteristics), it serves to benchmark how much money could be saved by a policy that can achieve this type of maximal-savings patient-physician matching.

To simulate such a policy, I begin with the dataset of all patients who switched physicians between 2015 and 2017. Within commuting zones, I re-assign patients to physicians based on their age, so that the middle-aged patients are assigned to the most cost-efficient physicians, the oldest patients are assigned to the next most cost-efficient physicians, and then the youngest patients are assigned to the least cost-efficient physicians. Each physician is assigned to the same number of patients that they treated in the original dataset, so capacity constraints are built into the exercise. I predict the change in outcomes from switching to the maximal-savings physicians by multiplying the age-specific effects in Table 9 by the changes in quality and cost efficiency between the original physician and the maximal-savings matched physician. After determining the potential savings from steering patients toward their maximal-savings matches, I also determine potential savings from steering patients randomly (within geographic areas).

I find that a policy that could steer patients toward their maximal savings matches could save 7.4% in annual spending over and above the savings from the matches patients endogenously made under the current policy. When compared to a policy that randomly allocates patients to physicians (which increases spending by 0.3%), the maximal savings matches save about 7.7%. While matches are not created based on health outcomes, the maximal-savings policy would also decrease preventable emergencies by about 2.7% of preventable visits relative to random matching (recall that the estimates on preventable emergencies were not statistically significant, so these estimates should be interpreted with caution). These numbers should be interpreted as order-of-magnitude estimates of what could occur under a policy that more effectively matches patients to physicians. However, these numbers should not be viewed as welfare estimates, since patient preferences over their physicians are explicitly ignored in this analysis.

This research draws several conclusions. First, privately insured patients respond to information about primary care physicians, and younger patients respond more. When patients do switch physicians, switching to more cost-efficient physicians decreases patients' spending without significantly impacting outcomes, as measured by preventable emergency room visits. Middle-aged patients are impacted the most by their physician's quality and cost efficiency, while the youngest patients are more heavily steered by the disclosure program to higher quality, more cost-efficient physicians. There is room for policy improvement through policies that focus information or financial incentives on middle-aged patients.

Works Cited

Alexander, Diane, Janet Currie, and Molly Schnell. 2019. "Check up before you check out: Retail clinics and emergency room use." *Journal of Public Economics* (Elsevier) 178: 104050.

Allende, Claudia, Francisco Gallego, and Christopher Neilson. 2019. "Approximating the equilibrium effects of informed school choice." *Working Paper*.

Anderson, Michael, and Jeremy Magruder. 2012. "Learning from the crowd: Regression discontinuity estimates of the effects of an online review database." *The Economic Journal* (Oxford University Press Oxford, UK) 122 (564): 957--989.

Beaulieu, Nancy Dean. 2002. "Quality information and consumer health plan choices." *Journal of Health Economics* (Elsevier) 21 (1): 43--63.

Bundorf, M Kate, Natalie Chun, Gopi Sha Goda, and Daniel P Kessler. 2009. "Do markets respond to quality information? The case of fertility clinics." *Journal of Health Economics* 28 (3): 718--727.

Calonico, Sebastian, Matias D Cattaneo, and Rocio Titiunik. 2014. "Robust nonparametric confidence intervals for regression-discontinuity designs." *Econometrica* 82 (6): 2295--2326.

Chan, David C, Matthew Gentzkow, and Chuan Yu. 2022. "Selection with variation in diagnostic skill: Evidence from radiologists." *The Quarterly Journal of Economics* (Oxford University Press) 137 (2): 729--783.

Chartock, Benjamin L. 2021. "Quality Discloure, Demand, and Congestion: Evidence from Physician Ratings." *Working Paper*.

Currie, Janet, and Bentley MacLeod. 2017. "Diagnosing Expertise: Human Capital, Decision Making, and Performance Among Physicians." *Journal of Labor Economics* (University of Chicago Press Chicago, IL) 35 (1): 1--43.

Currie, Janet, and Jonathan Zhang. 2021. "Doing More with Less: Predicting Primary Care Provider Effectiveness." *National Bureu of Economic Research*.

Currie, Janet, W Bently MacLeod, and Jessica Van Parys. 2016. "Provider practice style and patient health outcomes: The case of heart attacks." *Journal of Health Economics* (Elsevier) 47: 64--80.

Cutler, David M, Robers T Huckman, and Mary Beth Landrum. 2004. "The role of information in medical markets: an analysis of publicly reported outcomes in cardiac surgery." *American Economic Review* 94 (2): 342-346.

Dhalstrand, Amanda. 2021. "Defying Distance? The Provision of Services in the Digital Age." *Workign Paper*.

Doyle Jr, Joseph J, Steven M Ewer, and Todd H Wagender. 2010. "Returns fo physician human capital: Evidence from patients randomized to physician teams." *Journal of Health Economics* (Elsevier) 29 (6): 866-882.

Dranove, David and Jin, Ginger Zhe. 2010. "Quality Disclosure and Certification: Theory and Practice." *Journal of Economic Literature* 935--63.

Dranove, David, and Andrew Sfekas. 2008. "Start spreading the news: a structural estimate of the effects of New York hospial report cards." *Journal of Health Economics* (Elsevier) 27 (5): 1201-1207.

Dranove, David, and Ginger Zhe Jin. 2010. "Quality Disclosure and Certification: Theory and Practice." *Journal of Economic Literature* 48 (4): 935--63.

Dranove, David, Daniel Kessler, Mark McClellan, and Mark Satterthwaite. 2003. "Is more information better? The effects of "report cards" on health care providers." *Journal of Political Economy* 111 (3): 555--588.

Fadlon, Itzik, and Jessica Van Parys. 2020. "Primary care physician practice styles and patient care: Evidence from physician exits in Medicare." *Journal of Health Economics* (Elsevier) 71: 102304.

Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams. 2016. "Sources of geographic variation in health care: Evidence from patient migration." *The quarterly journal of economics* (MIT Press) 131 (4): 1681--1726.

Finkelstein, Amy, Sarah Taubman, Bill Wright, Mira Bernstein, Jonathan Gruber, Joseph P Newhouse, Heidi Allen, Katherine Baicker, and Oregon Health Study Group. 2019. "The Oregon health insurance experiment: evidence from the first year." *The Quarterly journal of economics* (MIT Press) 127 (3): 1057--1106.

Fletcher, Jason, Leora I Horwitz, and Elizabeth Bradley. 2014. "Estimating the value added of attending physicians on patietn outcomes." *National Bureau of Economic Research Working Paper*.

Goodman-Bacon, Andrew. 2021. "Difference-in-differences with variation in treatment timing." *Journal of Econometrics* (Elsevier) 225 (2): 254--277.

Gowrisankaran, Gautam, and Keither, Leger, Pierre Thomas Joiner. 2022. "Physician practice style and healthcare costs: Evidence from emergency departments." *Management Science* (INFORMS).

Grembi, Veronica, Tommaso Nannicini, and Ugo Troiano. 2016. "Do fiscal rules Matter?" *American Economic Journal: Applied Economics* (JSTOR) 1--30.

Gruber, Jonathan, and Robin McKnight. 2016. "Controlling health care costs through limited network insurance plans: Evidence from Massachusetts state employees." *American Economic Journal: Economic Policy* 8 (2): 219--50.

Grytten, Jostein, and Run Sorensen. 2003. "Practice variation and physician-specific effects." *Journal of Health Economics* (Elsevier) 22 (3): 403--418.

Herr, Annika, and Hanna Hottenrott. 2016. "Higher prices, higher quality? Evidence from German nursing homes." *Health Policy* (Elsevier) 120 (2): 179--189.

Hull, Peter. 2018. "Estimating treatment effects in mover designs." *arXiv preprint.*

Kwok, Jennifer. 2019. "How do primary care physicians influence healthcare? Evidence on practice styles and switchingcosts from Medicare." *Working Paper.*

Mayo, Jennifer. 2021. "Navigating the notches: charity responses to ratings." *Working Paper.*

McCrary, Justin. 2008. "Manipulation of the running variable in the regression discontinuity design: A density test." *Journal of econometrics* (Elsevier) 142 (2): 698--714.

Miller, Sarah. 2012. "The effect of insurance on emergency room visits: an analysis of the 2006 Massachusetts health reform." *Journal of Public Economics* (Elsevier) 96 (11-12): 893--908.

Molitor, David. 2018. "The evolution of physician practice styles: evidence from cardiologist migration." *American Economic Journal: Economic Policy* 10 (1): 326--56.

Penn, Madeline, Saurabha Bhatnagar, SreyRam Kuy, Steven Lieberman, Sheeref Elnahal, Carolyn Clancy, and David Shulkin. 2019. "Comparison of wait times for new patients between the private sector and the United States Department of Veterans Affairs Medical Centers." *JAMA Network Open* (American Medical Association) 2 (1): e187096--e187096.

Sabety, Adrienne. 2021. "The Value of Relationships in Healthcare." *Working Paper.*

Schnell, Molly. 2017. "Physician behavior in the presence of a secondary market: The case of prescription opioids." *Princeton University Department of Economics Working Paper.*

Schnell, Molly, and Janet Currie. 2018. "Addressing the opioid epidemic: is there a role for physician education?" *American Journal of Health Economics* (MIT Press) 4 (3): 383--410.

Simeonova, Emilia, Niels Skipper, and Peter R Thingholm. 2020. "Physician health management skills and patient outcomes." *National Bureau of Economic Research Working Paper.*

Van Parys, Jessica. 2016. "Variation in physician practice styles within and across emergency departments." *PloS one* (Public Library of Science San Francisco, CA USA) 11 (8): e0159882.

Vatter, Benjamin. 2021. "Quality disclosure and regulation: Scoring design in medicare advantage." *Working Paper.*

Wedig, Gerard J, and Ming Tai-Seale. 2002. "The effect of report cards on consumer choice in the health insurance market." *Journal of Health Economics* (Elsevier) 21 (6): 1031--1048.

Yoon, Tae Jung. 2020. "Qualiity information disclosure and patient reallocation in the healthcare industry: Evidence from cardiac surgery report cards." *Marketing Science* (INFORMS) 39 (3): 636--662.

Figures

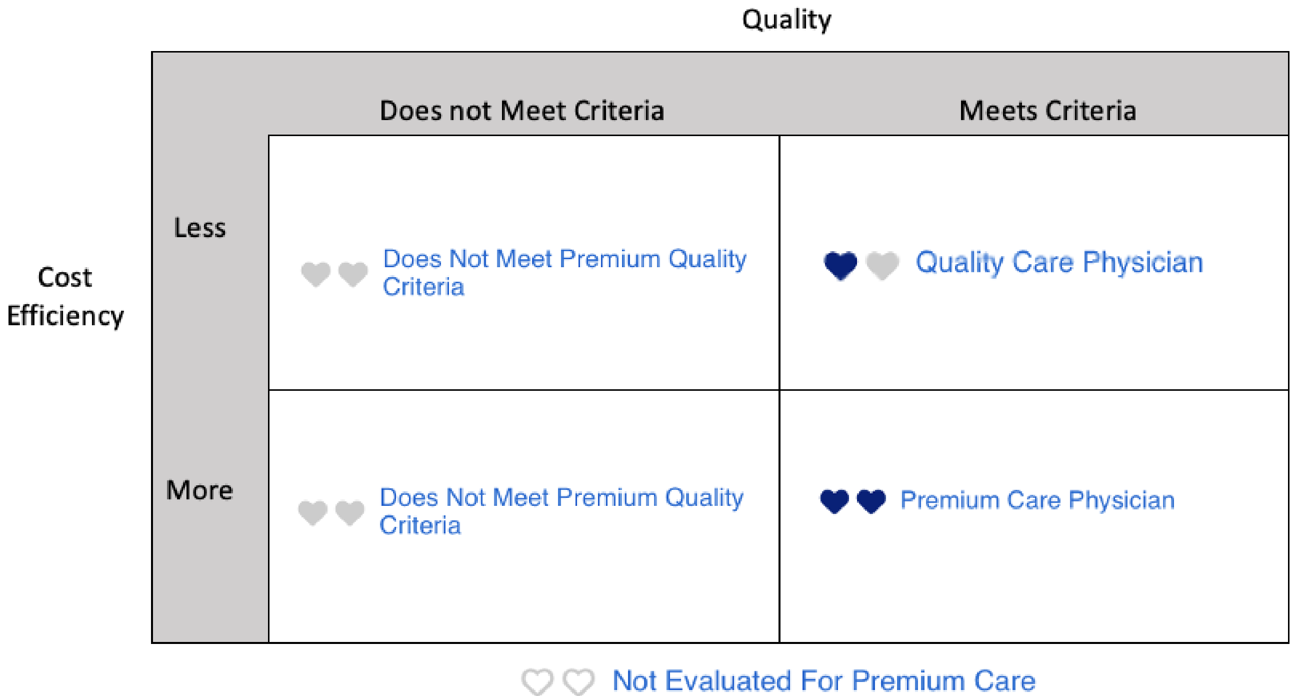
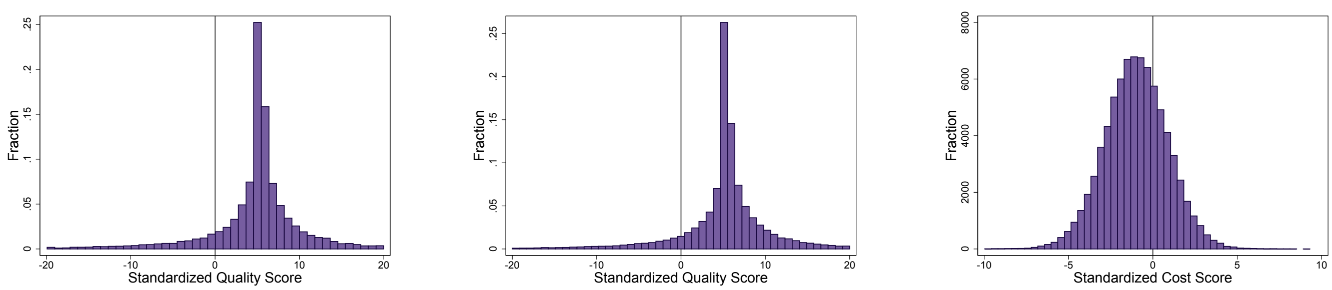


Figure 1: Physician Status as Displayed on Online Profile

Notes: This table displays designations by cost efficiency and quality status.



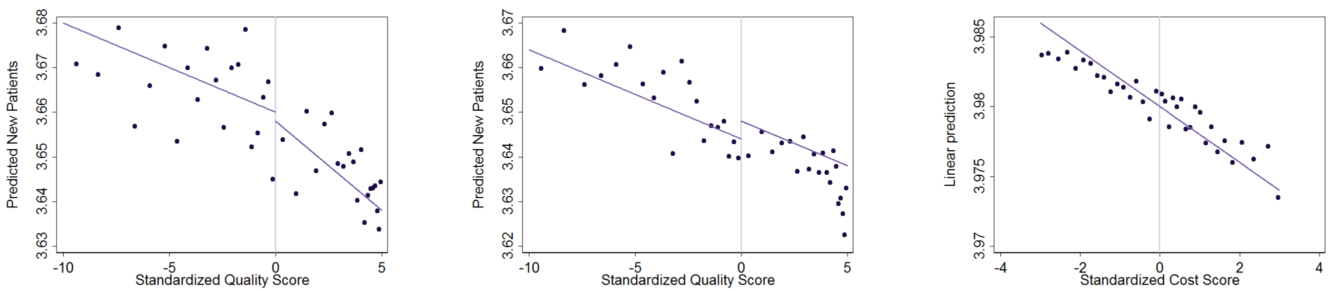
(a) Lower-cost Physicians Around Quality Cutoff: Zero vs. Two Hearts

(b) Expensive Physicians Around Quality Cutoff: Zero vs. One Hearts

(c) High Quality Physicians Around Cost Cutoff: One vs Two Hearts

Figure 2: Histograms

Notes: This figure displays three histograms. Panel A displays the histogram of underlying quality scores for cost efficient physicians. Panel B displays the same for non-cost efficient physicians, while panel C displays a histogram of underlying cost scores. There is no visual evidence of bunching around the cutoff.



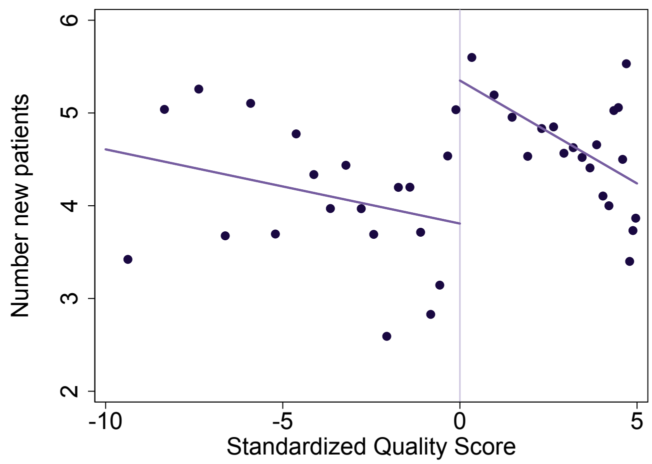
(a) Zero vs. Two Hearts

(b) Zero vs. One Heart

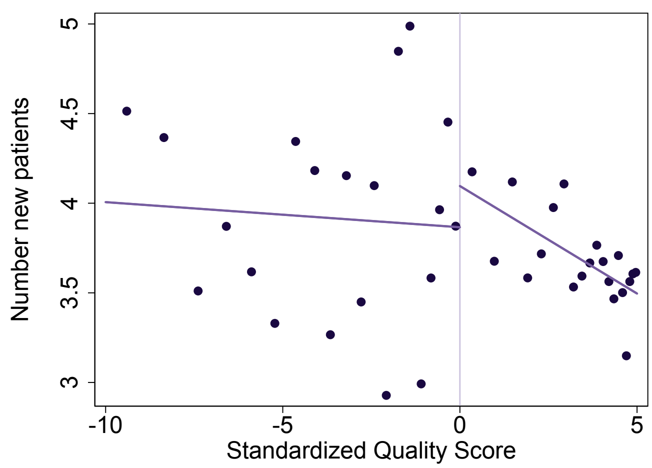
(c) One vs Two Hearts

Figure 3: Validity Test: Impacts on Predicted New Patients

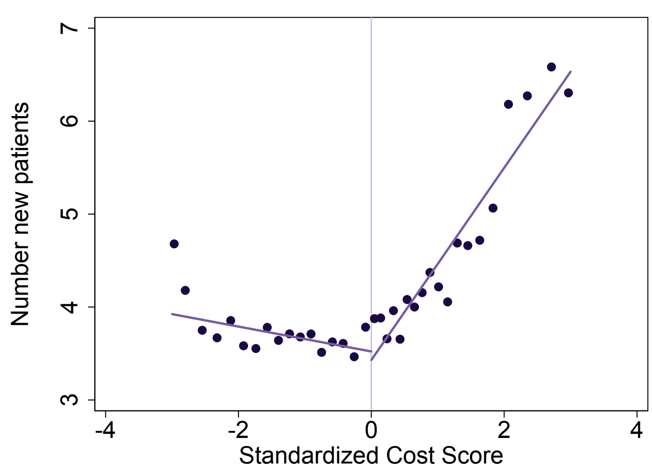
Notes: This figure displays three RD plots. All three display binned scatter plots of predicted number of new patients based solely on provider characteristics. If RD assumptions are satisfied, predicted new patients should be smooth around the cutoff. Panel A displays predicted patients around the quality threshold for more cost efficient physicians, who were assigned zero hearts if they missed the quality threshold, and two hearts if they achieved it. Panel B displays the same for less cost efficient physicians, who were only assigned one heart if they passed the threshold, and zero if they missed it. Panel C displays predicted values of new patients around the cost threshold for higher quality physicians. Those to the left of the cutoff were assigned one heart, while those to the right were assigned two. None of the figures show evidence of physician gaming or manipulating the threshold.



(a) Zero vs. Two Hearts



(b) Zero vs. One Heart



(c) One vs Two Hearts

Figure 4: These figures plot the number of new patients seen by providers over the three months following status updates. Outcomes are averaged in each bin for twenty quantile-spaced bins around the cutoff. Panel A shows the impact on new patients for cost efficient primary care physicians. These providers achieved zero hearts if below the cutoff, and two hearts if above. Panel B shows the impact on new patients for non-cost efficient physicians (comparing one heart to none), and Panel C shows the impact on new patients for high-quality physicians based on whether the physician passed the low-cost threshold (comparing two hearts to one).

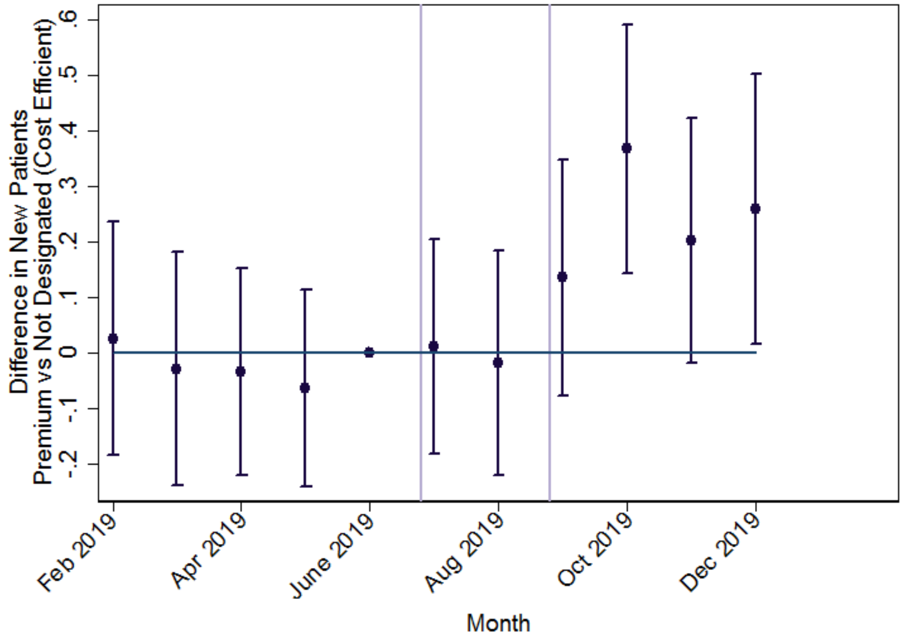
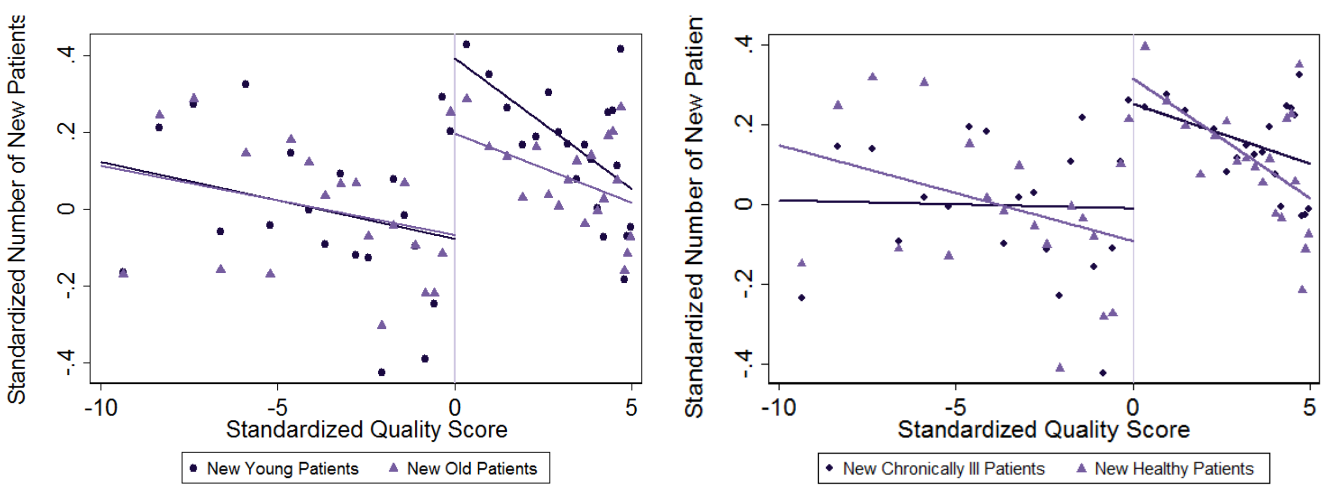


Figure 5: Difference in Discontinuities

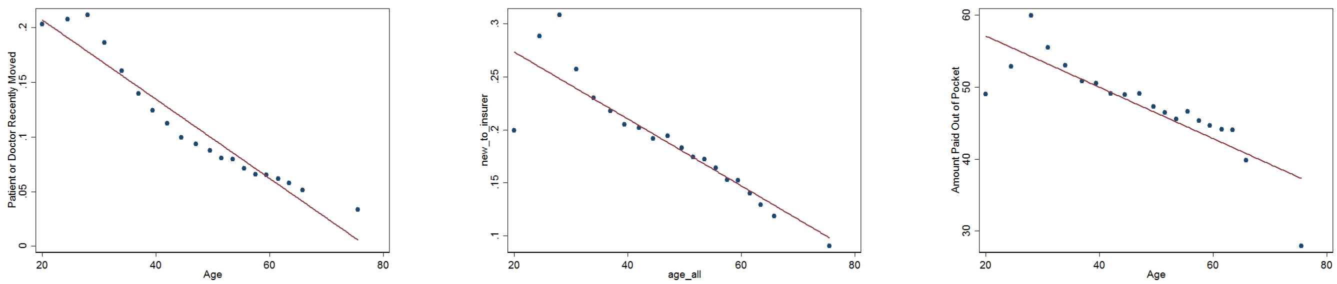
Notes: This figure shows β_k for months leading up to and following disclosure of updated statuses to providers (late June 2019) and patients (late August 2019) for low-cost primary care providers.



(a) Younger Versus Older Patients

(b) Chronically Ill Versus Healthy Patients

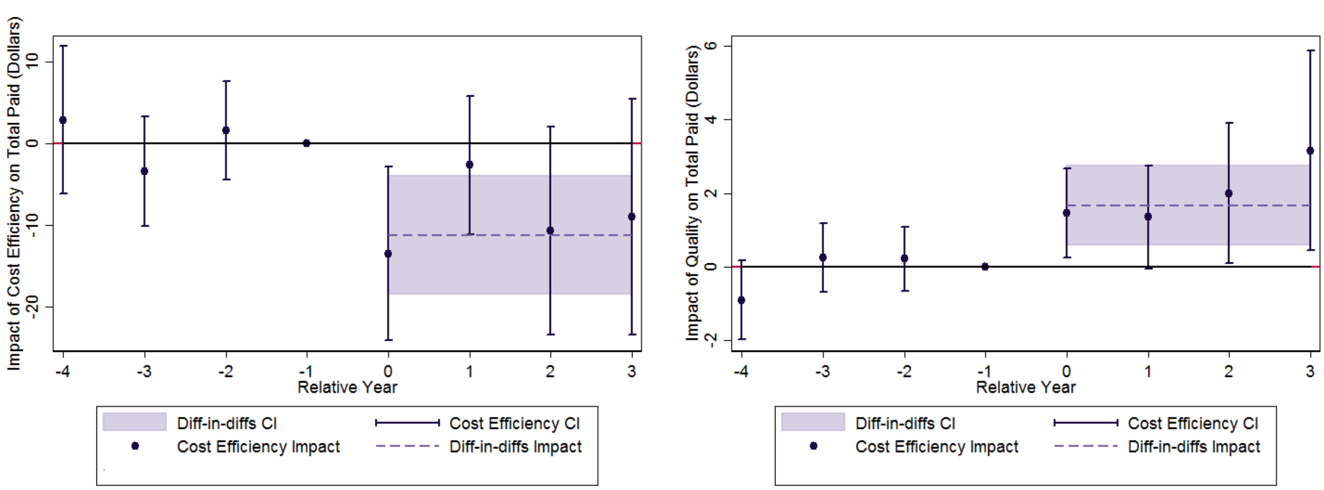
Figure 6: These figures plot the number of new patients seen by providers over the three months following status updates for lower-cost primary care providers. New patients are broken down into older (age 40 and above) and younger (age 18-39) patients in panel A and into chronically ill versus healthy patients in panel B. Outcomes are averaged in each bin for twenty quantile-spaced bins around the cutoff.



(a) Recently Moved (b) New to Insurer (c) Out of Pocket Payment

Figure 7: Binned Scatterplots of Age Against Search Characteristics

Notes: These figures display binned scatterplots for patient characteristics against age. Younger patients are more likely to have recently moved or for their physician to have recently moved, are more likely to be new to the insurer, and face higher out of pocket payments for their upcoming visit.



(a) Effect of Cost Efficiency (b) Effect of Quality

Figure 8: Effect of quality and cost efficiency on total amount paid

Notes: This figure displays the impacts of switching to a single point more cost efficient (panel A) or higher quality (panel B) physician on the total amount paid: the sum of patient- and insurer spending.

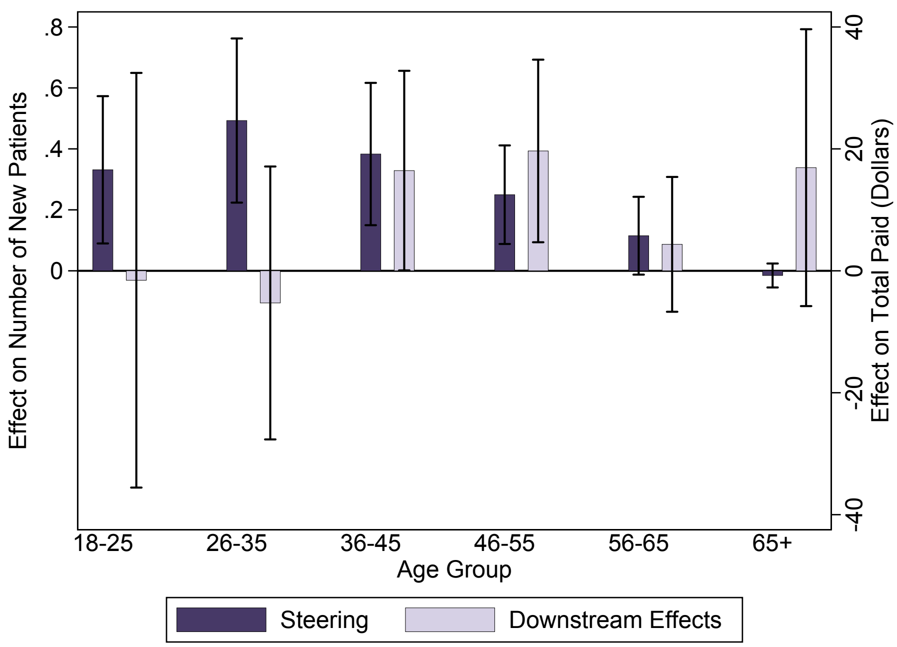


Figure 9: Comparison of Steering to Downstream Outcome Effects by Age Group

Notes: This figure plots, in dark purple, impacts of the Premium designation on the number of new patients in each age group. In light purple, the impact of seeing a one-point more cost efficient physician is displayed separately for each age group. The patients who respond most strongly do the disclosure policy are those who benefit the least from switching.

Tables

	Not Designated	Quality	Premium	All
Quality Score	–17.96	1.58	1.74	–0.66
	(27.20)	(5.50)	(6.64)	(12.60)
Cost Score	–0.46	0.82	–1.66	–0.30
	(1.84)	(1.40)	(1.42)	(1.88)
Total Patients	36.56	38.46	40.14	38.90
	(60.78)	(164.56)	(51.70)	(121.48)
New Patients	4.30	4.00	4.38	4.20
	(10.60)	(7.82)	(8.06)	(8.30)
Older New Patients	2.00	2.08	2.10	2.08
	(4.80)	(3.72)	(3.58)	(3.82)
Younger New Patients	2.32	1.94	2.28	2.12
	(6.50)	(5.08)	(5.30)	(5.36)
Chronically Ill New Patients	0.96	0.92	1.02	0.96
	(2.24)	(1.84)	(1.94)	(1.92)
Non-Chronically Ill New Patients	3.36	3.08	3.38	3.22
	(8.74)	(6.44)	(6.60)	(6.82)
Observations	10,616	44,347	35,519	90,482
Proportion of Total	0.12	0.50	0.40	1.00

Table 1: Summary Statistics

Notes: This table reports summary statistics from claims with a date between zero and three months of public disclosure of provider status, broken down by provider status. The first two rows summarize underlying continuous cost and quality scores, and the remainer of the rows summarize patient volume. Older new patients are those patients aged 40 and older who saw their doctor for the first time, while younger new patients are those aged 18-39 who saw their doctor for the first time. Chronically ill patients are those who have a diagnosis of any chronic illness from an office visit over the past year. Proportions as displayed are rounded, and rounding error explains why the proportions do not sum to one.

	(1)	(2)	(3)	(4)
	Female	Cost Score	Graduation Year	Healthgrades Score
Panel A: Low-Cost Doctors 0 vs 2 Hearts				
Above cutoff	0.0130	0.0228	-0.903	-0.142
	(0.0211)	(0.0472)	(0.780)	(0.129)
Observations	8743	8743	4305	843
Panel B: Expensive Doctors 0 vs 1 Hearts				
Above cutoff	-0.0172	-0.0767*	-0.733	0.0590
	(0.0151)	(0.0411)	(0.548)	(0.106)
Observations	19403	19403	9027	1999
Panel C: Cost Threshold 1 vs 2 Hearts				
Below Cost Cutoff	0.00484	0.140	-0.0526	0.0315
	(0.00714)	(0.0919)	(0.242)	(0.0390)
Observations	67644	67644	30545	7707

Table 2: Covariate Continuity

Notes: This table reports results of the regression in equation 1 on physician characteristics as outcomes.

	(1)	(2)	(3)	(4)
	Number new patients	Number new patients	Number new patients	Number new patients
High Quality=1	0.642***	1.543***	0.230	
	(0.239)	(0.437)	(0.286)	
Cost Efficient=1				-0.0921
				(0.110)
Heterogeneity Variable	Primary Care	Primary Care	Primary Care	Primary Care
Cost or Quality Status	All	Cost Efficient	Not Cost Efficient	High Quality
R^2	0.000592	0.00187	0.000553	0.00602
Observations	28146	8743	19403	67644

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Main Effects for Primary Care Providers

Notes: This table displays main effects for primary care providers. Column 1 displays pooled results over cost status, while columns 2 and 3 show results for cost efficient and non-cost efficient physicians separately. Column 4 displays the impact of passing the cost threshold for the higher quality providers who met the quality criteria.

	(1)	(2)	(3)
	Age	Chronic Illness	Last Year's Spending
Above Cutoff	-1.407**	-0.0243*	-80.61
	(0.613)	(0.0127)	(170.7)
Cost Status			
Observations	39176	39176	39176

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Patient Heterogeneity

Notes: This table displays the regression discontinuity estimation with patient characteristics as outcomes for cost efficient primary care physicians. Column 1 displays the impact of the Premium designation on the average age of new patients seen, and columns 2 and 3 report impacts on chronic illness status and the patient's prior year total spending (insurer + patient). The designation itself does not change patient characteristics, instead, these impacts should be interpreted as resulting from different responses of patients to Premium status based on the patient's characteristics.

	(1)	(2)	(3)
	Patient or Doctor Moved	New to Insurer	Out of Pocket
Above Cutoff	-0.00598	0.00173	10.44**
	(0.00934)	(0.0130)	(5.283)
Cost Status			
Observations	39176	39176	39174

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Patient Search Mechanisms

Notes: This table displays the results of individual level regressions estimating equation 1 on characteristics related to patient search: whether the patient or physician recently moved, whether the patient is new to the insurer, and the patient's out of pocket payment.

	Stayers	Switchers	Induced
Total Paid	715.248	681.086	679.252
	778.778	1125.8	744.71
Services per Visit	2.27	2.206	2.204
	1.778	1.762	1.736
Services per Visit	2.27	2.206	2.204
	1.778	1.762	1.736
Price per service	131.986	145.68	142.78
	90.418	109.682	102.97
Total Paid All	5031.834	5562.312	5503.856
	13963.19	15242.05	15812.18
Non-modal Doctor Paid	4316.584	4881.226	4824.604
	13856.5	15096.22	15713.39
Preventable ED	.022	.02	.026
	.506	.45	.51
Non-Emergency ED Visits	.006	.008	.006
	.164	.156	.114
Placebo ED Visits	.002	.002	.002
	.252	.242	.162
ΔQ	0	-.868	-.45
	0	17.494	16.108
ΔC	0	-.086	-.256
	0	2.82	2.682
ΔQC	0	.082	-.726
	0	46.142	40.104
Quality Score	2.408	1.872	1.274
	12.86	13.35	12.252
Cost Efficiency Score	.398	.278	.35
	2.114	2.102	2.022
Observations	1343489	632450	41972

Table 6: Downstream Outcomes Summary Statistics

Notes This table displays summary statistics for the dataset used in downstream outcomes analysis. The average and standard deviation are displayed for each outcome variable and for right hand side variables. Patients are broken into three groups. Stayers are those patients who see the same primary care physician over the duration of the panel. Switchers are those who switch physicians, and Induced patients are those whose switch was induced by their former PCP either leaving or moving.

	(1) Total Paid	(2) Services	(3) Services per Visit	(4) Price per service	(5) Total Paid All	(6) Non-modal Doctor Paid	(7) Patient Paid
$\Delta Q \times$ Post	1.664***	0.0179***	0.00546***	-0.109	16.45	14.79	0.312
	(0.547)	(0.00666)	(0.00156)	(0.0753)	(11.03)	(11.02)	(0.233)
$\Delta C \times$ Post	-11.18***	-0.0107	0.0104	-1.612***	-108.2**	-97.04**	-3.031*
	(3.670)	(0.0410)	(0.00816)	(0.515)	(44.78)	(44.49)	(1.737)
R-Squared	0.415	0.509	0.505	0.546	0.453	0.451	0.454
Outcome Mean	679.26	6.32	2.2	142.78	5503.86	4824.6	185.34
Average Impact	-16.56	.02	.04	-3.14	-159.56	-143	-4.78
Observations	41972	41972	41972	41972	41972	41972	41972

Standard errors in parentheses

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

Table 7: Effects of Quality and Cost Efficiency on Spending Outcomes

Notes: This table shows the impacts of quality and cost efficiency on spending outcomes. Column 1 reports the impacts on total spending: the amount paid for visits with the modal primary care physician by both the patient and the insurer. Column 2 reports impacts on the number of service received by the modal primary care physician. The outcome in column 3 is the average number of services received per visit, and column 4 reports the average price per service: total amount paid divided by total number of services. Columns 5 and 6 explore possible spillovers in spending on other providers. The outcome for the regression in column 5 is spending over all providers, not just the modal primary care provider. Column 6 narrows down to spending over other providers, not including the modal primary care provider. Column 7 displays impacts on the portion the patient paid out of pocket.

	(1) Preventable ED	(2) Unnecessary ED	(3) Placebo ED
$\Delta Q \times$ Post	-0.000323	0.000248	-0.0000880
	(0.000690)	(0.000208)	(0.000365)
$\Delta C \times$ Post	0.0000734	0.000977	0.00226*
	(0.00338)	(0.000948)	(0.00124)
R-Squared	0.311	0.190	0.149
Outcome Mean	.083	.02262	.02342
Average Impact	-.00056	.00228	.00388
Observations	41972	41972	41972

Standard errors in parentheses

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

Table 8: Effects of Quality and Cost Efficiency on Emergency Department Utilization

Notes: This table shows the impacts of quality and cost efficiency on emergency room utilization outcomes. Column 1 displays the impacts on preventable emergency department (ED) visits, which are conditions that are true emergencies but which appropriate primary care could have in part prevented. Column 2 explores impacts on non-emergency ED visits, which are non-emergencies that would more appropriately be treated in a primary care office setting. Column 3 reports impacts on "placebo" emergency department visits, which are true emergencies that are not preventable by higher quality or more cost efficient preventative care.

	(1)	(2)	(3)	(4)	(5)	(6)
	Total Paid	Total Paid	Total Paid	Total Paid	Total Paid	Total Paid
$\Delta Q\times$ Post	5.493	2.465	3.920**	2.214**	0.590	2.929
	(4.086)	(3.407)	(1.584)	(0.955)	(0.667)	(2.946)
$\Delta C\times$ Post	1.549	5.273	-16.44**	-19.66**	-4.342	-16.92
	(17.13)	(11.36)	(8.336)	(7.635)	(5.639)	(11.57)
Age Range	18-25	26-35	36-45	46-55	56-64	65 and older
Outcome Average	635.68	610.7	660.96	690.9	727.6	560.62
Average Impact	.28	13.2	-16.14	-29.64	-7.06	-28.9
Standard Error	(26.6)	(17.56)	(13.22)	(12.04)	(8.76)	(17.94)
R-Squared	0.432	0.437	0.435	0.366	0.446	0.413
Observations	658	1463	4648	11445	16121	6174

Standard errors in parentheses

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

Table 9: Effects on Total Spending By Age Group

Notes: This table shows the impacts of quality and cost efficiency on total spending, broken into different age groups. The impacts for the youngest adult patients, aged 18-25 are shown in column 1, and age increases over the columns with the oldest patients, aged 65 and above, shown in column 6.

	(1)	(2)	(3)	(4)	(5)	(6)
	Preventable ED	Preventable ED	Preventable ED	Preventable ED	Preventable ED	Preventable ED
$\Delta Q\times$ Post	-0.000326	-0.00304	0.00363	-0.00158	-0.000111	0.00198
	(0.00248)	(0.00354)	(0.00353)	(0.00126)	(0.000533)	(0.00125)
$\Delta C\times$ Post	0.0756	-0.0487	-0.00168	0.0140***	-0.00724*	-0.00635
	(0.0770)	(0.0436)	(0.00483)	(0.00505)	(0.00386)	(0.00723)
Age Range	18-25	26-35	36-45	46-55	56-64	65 and older
Outcome Average	.08	.04	.02	.02	.02	.02
Average Impact	.16	-.11	.006	.026	-.016	-.008
Standard Error	(.12)	(.068)	(.008)	(.008)	(.006)	(.012)
R-Squared	0.287	0.189	0.302	0.361	0.295	0.408
Observations	658	1463	4648	11445	16121	6174

Standard errors in parentheses

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

Table 10: Effects on Preventable Emergencies By Age Group

Notes: This table shows the impacts of quality and cost efficiency on preventable emergency department visits, broken into different age groups. The impacts for the youngest adult patients, aged 18-25 are shown in column 1, and age increases over the columns with the oldest patients, aged 65 and above, shown in column 6.

	(1)	(2)	(3)	(4)	(5)	(6)
	Total Paid	Total Paid	Total Paid	Total Paid	Total Paid	Total Paid
$\Delta Q\times$ Post	9.601**	-0.918	0.763	4.280***	1.807	0.521
	(3.988)	(3.360)	(0.956)	(1.057)	(1.307)	(0.889)
$\Delta C\times$ Post	17.78	-4.011	-20.44	-18.13***	-9.788	-7.299
	(11.63)	(13.43)	(14.50)	(5.184)	(9.452)	(5.968)
Age Range	18-35	18-35	36-55	36-55	56 and older	56 and older
Health Status	No Health Change	Newly Chronically Ill	No Health Change	Newly Chronically Ill	No Health Change	Newly Chronically Ill
R-Squared	0.459	0.413	0.450	0.357	0.400	0.480
Observations	1078	1043	5509	10584	8281	14014

Standard errors in parentheses

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

Table 11: Effects on Preventable Emergencies By Age Group

Notes: This table shows the impacts of quality and cost efficiency on preventable emergency department visits, broken into different age and chronic illness groups. The impacts for the youngest adult patients are in columns 1 and 2, and age increases over the columns with the oldest patients shown in columns 5 and 6. The columns also break down patients by chronic illness status, with columns 1, 3, and 5 displaying impacts for patients without any change in chronic illness status over the time period studied, and columns 2, 4, and 6 displaying impacts for those with a new or recent chronic illness diagnosis.

Appendix

A.1 Difference between impacts of achieving Quality versus Premium designation

The main results of this study show that patients respond more to the Premium Care designation (two hearts on the physician's profile) versus the Quality Care designation (one heart on the physician's profile). This section discusses possible reasons for the differences.

First, one may wonder whether patients correctly interpret the second heart shown on the profile as a measure of cost efficiency rather than an indicator of even higher quality. Perhaps the difference in effects is driven by patients responding to what they perceive as a measure of "very high quality" rather than "high quality and cost-efficient." To determine whether this is the case, I examined the impacts of heart status separately for patients who paid nothing out-of-pocket for their visit. If these patients know ahead of time that they stand to pay nothing for their visit, they should ignore the second heart since the information about cost efficiency it provides is irrelevant. If premium (two-heart) physicians are still prioritized over single-heart physicians, it is either because patients misunderstand the meaning of the second heart, or because the response is driven by page rank rather than responses to the information provided. Appendix Figure 9 displays the results of this exercise. Panel A compares the responses of patients who paid nothing for their visit to those who paid for 100% of their visit out-of-pocket for cost-efficient physicians (comparing two-heart Premium physicians to zero-heart Not Designated physicians). My findings show clear effects for both types of patients. Panel B compares the same for non-cost-efficient physicians and shows no effect. These results suggest that either patients misinterpret the information provided by the hearts (thinking that premium physicians are even higher quality rather than high quality and cost-efficient) or that the impact of page-rank drives the results.

Second, one may hypothesize that regression discontinuity estimates are local, which drives the differences between one-heart and two-heart physicians. Perhaps patients prefer single-heart Quality physicians over double-heart Premium physicians, but that single-heart Quality physicians have a lower capacity to take on new patients. This would make treatment effects are mechanically lower. Appendix Figure 10 shows that providers who are not at capacity (as proxied by the provider taking on no new patients but having seen a positive number of returning patients during the pre-update period) experience large treatment effects, whereas physicians who are at capacity experience none. However, there were no significant differences in the proportion of physicians who were at capacity between these groups. The impact difference between cost-efficient and non-cost-efficient providers remains when subsetting down to physicians who are not at capacity (see Appendix Figure 10). Capacity constraints, therefore, do not explain the results.

A.2 Treatment Time Heterogeneity

Recent advances in the econometrics literature have pointed to homogeneity assumptions which can prove critical in two-way fixed effects designs such as the switcher design. Goodman-Bacon (2021) shows that the estimate of interest in a two-way fixed effect design estimates a weighted combination of treatment effects relative to various controls, where weights may have different signs. To overcome this estimation problem, researchers propose estimators which explicitly assign control groups to avoid bias caused by the aggregation implicit in OLS estimates of these two-way fixed effects designs (Callaway and Sant'Anna 2021; Sun and Abraham 2021).

Novel estimators require a very specific setup, where individuals are compared to each other pre- versus post-treatment, so the specification in Equation 9 cannot be used to estimate impacts that are robust to treatment time heterogeneity because that specification requires the post-treatment indicator to interact with the change in quality and cost efficiency.

In lieu of estimating robust to treatment time heterogeneity (Eq. 9), which is not possible given currently available estimators, I use Callaway and Sant'Anna's (2021) method to estimate the following specification:

$$Y_{it} = \iota_i + \tau_t + \beta_1 1\{PostSwitch\}_{it} + \nu_{it} \quad (11)$$

The regression in Equation 11 estimates the impact of switching physicians, regardless of the direction of the switch. To estimate impacts that are more comparable to the results of estimating the regression in Equation 9, I subset to groups of patients based on the direction of their switch. I first subset to patients who switched to a physician with a higher quality level, regardless of cost efficiency. Second, I subset to patients that switched to a physician with a higher cost-efficiency level, regardless of quality. Next, I subset based on both quality and cost efficiency, breaking patients down into four groups: those who switched to higher quality, lower cost-efficiency physicians, those who switched to physicians that were higher on both quality and cost efficiency, those who switched to physicians that were lower on both quality and cost efficiency, and those who switched to lower quality, higher cost efficiency physicians.

The regression in Equation 11 can be estimated using both OLS and the Callaway and Sant'Anna (CS) estimators. Comparing the two estimation methods can provide information on whether heterogeneous effects based on treatment time are biasing the OLS results. Breaking down patients into different categories helps to evaluate whether, qualitatively, patterns match those discussed above.

Appendix Figure 11 displays the results. The darker purple bars show the estimates of OLS regressions, while the lighter purple bars display the results of the analogous CS regression. Across the board, the CS estimates are larger than OLS estimates, suggesting that in this setting, OLS may bias impacts toward zero. The second point to notice is that the impact of switching a physician is always positive. That is, switching physicians results in higher spending, regardless of the cost efficiency or quality of the physician one is switching to. This result is also observed in the above estimation of Equation 11; however, the positive impact of switching on spending is captured by the relative year-fixed effects in x_{it}. Appendix Table 13 displays the full regression results for the specifications with relative year-fixed effects and with post-switch indicators separately.

Qualitatively, the OLS and CS impacts look quite similar. First, consider a patient who switches to a higher quality, less cost-efficient physician. If the results in Section 9.1 are true, then this should be the most expensive switch because cost efficiency declines while quality (which leads to higher spending) increases. This is also true as shown in Appendix Figure 11, where the bars for switching to higher quality, lower cost-efficiency physicians are the highest for both OLS and CS. On the other hand, a switch to a lower quality and higher cost-efficiency physician should lead to the smallest increase in spending. Indeed, as seen in Appendix Figure 11, bars for these switches (shown at the far right of the figure) are the lowest.

While estimation techniques to allow for robust estimation of the two-way fixed-effects design in this study are still being finalized in the literature, these checks lead me to conclude that treatment time heterogeneity may attenuate my OLS estimates while preserving the relative impacts of patient switches across physicians of various quality and cost efficiency.

A.3 Placebo Hospital Outcomes

Table 8 shows a marginally statistically significant increase in "placebo" emergency room visits—any visit to the emergency room for childbirth, fractures, or poisonings. It is important to understand whether this is a true, robust result. If it is, then one might be concerned that the patient-fixed effects were not absorbing all the essential patient-specific variations. For example, if the patients switch to higher quality or more cost-efficient providers upon finding they are pregnant (a time-varying characteristic), then the fixed patient effect would not absorb this type of selection bias. This would be concerning because impacts on total spending would also be biased by this issue.

I explore robustness in three ways. First, I break down placebo emergency visits into three components to determine if any one of the three (for example, childbirth) drives effects (see Appendix Table 6). There were no statistically significant impacts on any of the three outcomes. Second, I subset to the sample, which is balanced on time relative to the switch, to ensure that any impacts were not driven by sample imbalance. Appendix Table 7 displays these results, where again, my findings show no evidence of a statistically significant impact on any of the emergency department visit outcomes. Finally, to ensure the result is not simply under-powered, I display the results from an analysis of all switches – not just induced switches – in Appendix Table 8 and again find no evidence of impacts. Thus, I conclude that the marginal increase in placebo emergencies is spurious.

Works Cited

Callaway, Brantly, and Pedro H. C. Sant'Anna. 2021. "Difference-in-differences with multiple time periods." *Journal of Econometrics* 225 (2): 200-230.

Goodman-Bacon, Andrew. 2021. "Difference-in-differences with variation in treatment timing." *Journal of Econometrics* (Elsevier) 225 (2): 254-277.

Sun, Liyang, and Sarah Abraham. 2021. "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects." *Journal of Econometrics* (Elsevier) 225 (2): 175-199.

Appendix Figures and Tables

CPT Code	Description
99201	New patient outpatient visit, low complexity, low severity
99202	New patient outpatient visit, low complexity, low to moderate severity
99203	New patient outpatient visit, low complexity, moderate severity
99204	New patient outpatient visit, moderate complexity, moderate to high severity
99205	New patient outpatient visit, high complexity, moderate to high severity
99381	Initial comprehensive preventative medicine evaluation for a new patient: infant
99382	Initial comprehensive preventative medicine evaluation for a new patient: 1-4 years
99383	Initial comprehensive preventative medicine evaluation for a new patient: 5-11 years
99384	Initial comprehensive preventative medicine evaluation for a new patient: 12-17 years
99385	Initial comprehensive preventative medicine evaluation for a new patient: 18-39 years
99386	Initial comprehensive preventative medicine evaluation for a new patient: 40-46 years
99387	Initial comprehensive preventative medicine evaluation for a new patient: 65 years and older
92004	Opthalomogical services: new patient
92002	Opthalomogical services: new patient with diagnostic treatment program

Table 1: Procedure Codes for New Patients

Notes: This table lists the set of procedure codes used to identify new patient visits.

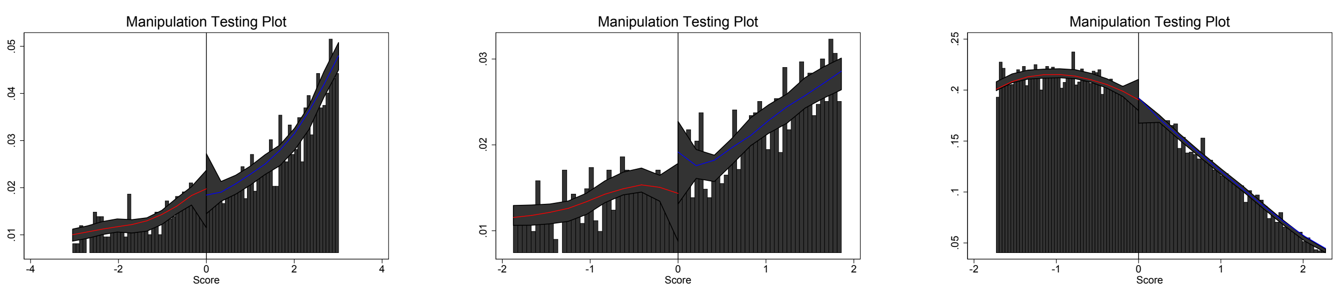
	(1)
	Number new patients
Female	-0.215
	(-0.54)
Constant	3.712***
	(20.15)
Observations	1587
R-Squared	0.489

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2: Predicting New Patients Using Provider Characteristics

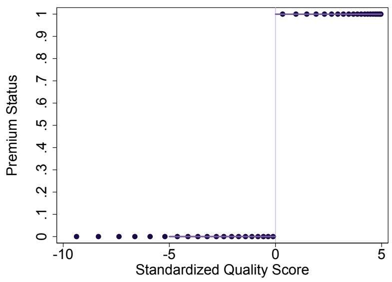
Notes: This table shows the results of a regression which predicts new patients based on physician gender with ZIP code fixed effects. The table uses a cross-section of primary care provider visits from the three months before providers were notified of their new scores.



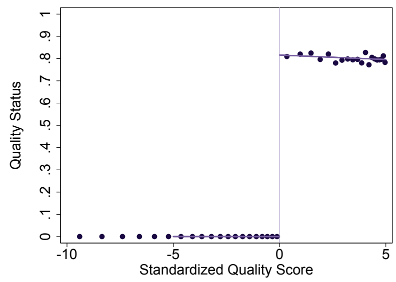
(a) Zero vs. Two Hearts (b) Zero vs. One Heart (c) One vs Two Hearts

Figure 1: Density Tests

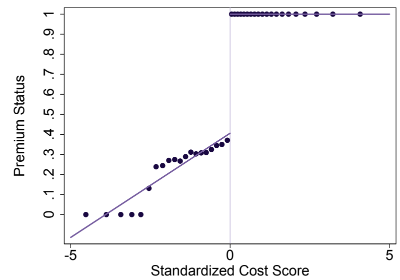
Notes: These figures display the output of density testing from Calonico et al. (2014)'s procedure. There are no statistically significant discontinuities in density across the three designs.



(a) Zero vs. Two Hearts



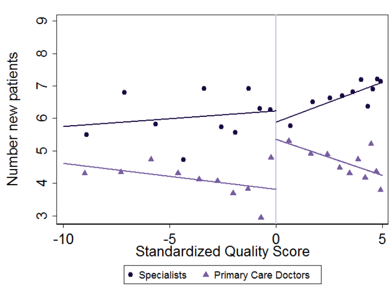
(b) Zero vs. One Heart



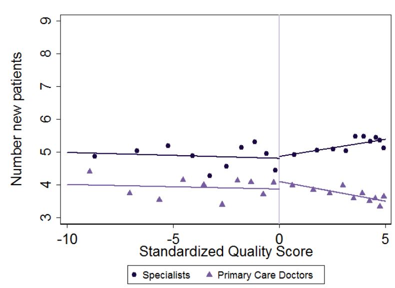
(c) One vs Two Hearts

Figure 2: First Stage

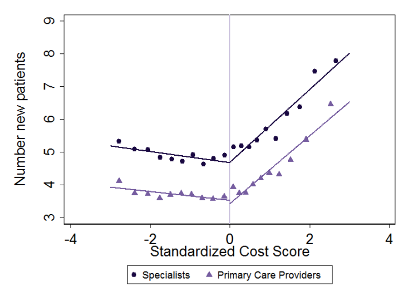
Notes: The proportion of Quality or Premium status physicians in each bin for twenty quantile-spaced bins around the cutoff is plotted over bins in Panels A-C. Panel A shows the first stage impact on status for more cost efficient physicians, panel B shows the first stage for less cost efficient physicians, Panel C shows the impact for higher-quality physicians. Compliance is not perfect in panels B and C since some physicians who would have had quality status are bumped up to premium status if their practice group is cost efficient, even if they individually are not. This is only the case for cost efficiency: physicians who do not meet quality criteria are always classified as Not Designated regardless of their group's behavior.



(a) Zero vs. Two Hearts



(b) Zero vs. One Heart



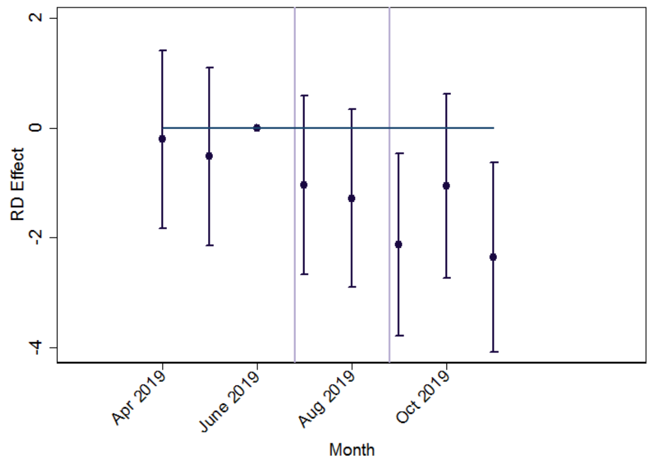
(c) One vs Two Hearts

Figure 3: These figures plot the number of new patients seen by providers over the three months following status updates separately for primary care providers and specialists. Outcomes are averaged in each bin for twenty quantile-spaced bins around the cutoff. Panel A shows the impact on new patients for cost efficient primary care physicians. These providers achieved zero hearts if below the cutoff, and two hearts if above. Panel B shows the impact on new patients for non-cost efficient physicians (comparing one heart to none), and Panel C shows the impact on new patients for high-quality physicians based on whether the physician passed the low-cost threshold (comparing two hearts to one).

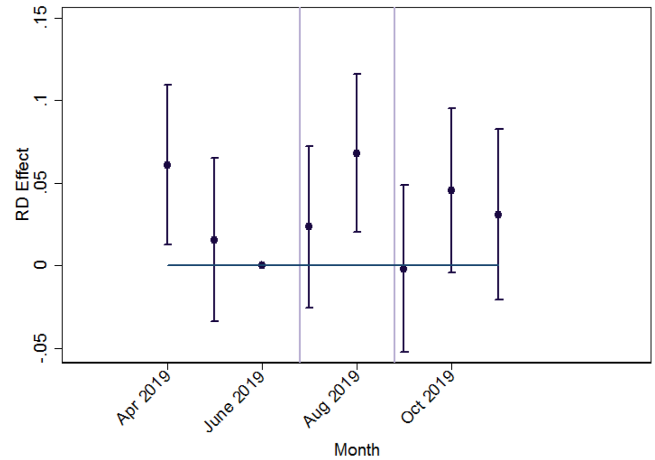
	(1)	(2)	(3)
	Quality Status	Number new patients	Number new patients
Panel A: Quality Fuzzy RD			
Above Quality Cutoff	0.815***	0.230	
	(0.0113)	(0.286)	
Quality Status			0.282
			(0.351)
Regression Type	First Stage	Reduced Form	IV Regression
F	6446.9	3.722	
Observations	19403	19403	19403
R^2	0.499	0.000553	0.000469
	(1)	(2)	(3)
Panel B: Cost Fuzzy RD			
Above Cost Cutoff	0.660***	0.399	
	(0.0141)	(0.306)	
Premium Status			0.605
			(0.464)
Regression Type	First Stage	Reduced Form	IV Regression
F	.	3.486	
N	7778	7778	7778
R^2	0.514	0.00139	0.000280

Table 3: IV Estimates

Notes: This table displays instrumental variables estimates from a fuzzy RD design where crossing the threshold is an instrument for having quality or premium status.



(a) Age of New Patients

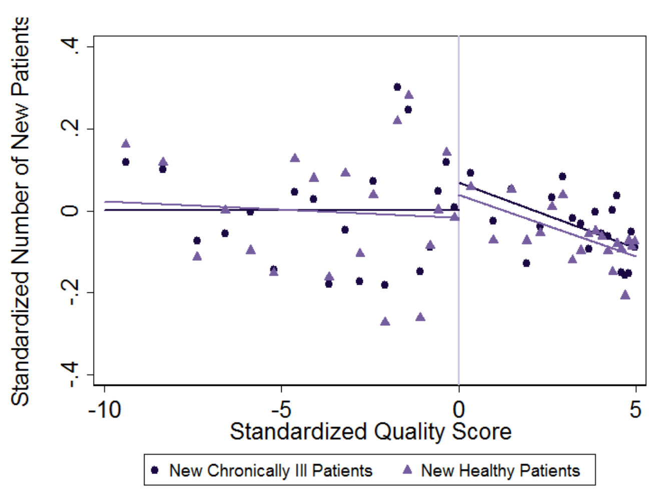


(b) Prop. Chronically Ill of New Patients

Figure 4: Difference in Discontinuity on Patient Characteristics

Notes: These figures plot the average age and proportion chronically ill new patients of providers over the three months following status updates for lower-cost primary care providers. Panel A displays average age of new patients, and Panel B displays proportion chronically ill. Outcomes are averaged in each bin for twenty quantile-spaced bins around the cutoff.

(a) Younger Versus Older Patients



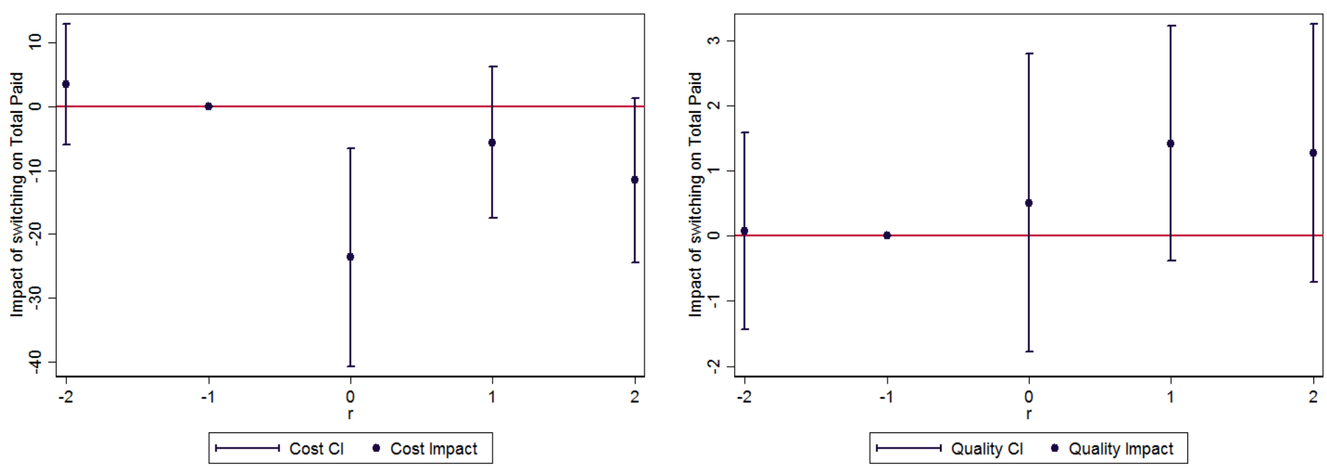
(b) Chronically Ill Versus Healthy Patients

Figure 5: These figures plot the number of new patients seen by more expensive providers over the three months following status updates for primary care providers. New patients are broken down into older (age 40 and above) and younger (age 18-39) patients in panel A and into chronically ill versus healthy patients in panel B. Outcomes are averaged in each bin for twenty quantile-spaced bins around the cutoff.

Dataset	Number of unique patients
Patients who have at least one medical claim each year	4,699,468
Patients whose modal primary care physician is eligible for a quality/cost efficiency score	2,657,447
Patients who have only one modal primary care physician	2,618,109
Patients who see a primary care physician each year	197,923
Patients with induced switches	41,972

Table 4: Switcher Analysis Dataset Sampling

Notes: This table displays the number of unique patients in various subsets of claims data used in the switcher analysis. Subsetting to patients who interact with the medical system once per year removes a large portion of patients. Other subsetting decisions are more marginal.



(a) Effect of Cost Efficiency

(b) Effect of Quality

Figure 6: Effect of quality and cost efficiency on total amount paid

Notes: This figure displays the impacts of switching to a single point more cost efficient (panel A) or higher quality (panel B) physician on the total amount paid: the sum of patient- and insurer spending estimated on a panel which is balanced on years relative to the switch year.

	(1) Total Paid	(2) Services	(3) Services per Visit	(4) Price per service	(5) Total Paid All	(6) Non-modal Doctor Paid	(7) Patient Paid
$\Delta Q\times$ Post	1.623**	0.0177**	0.00514***	-0.154*	16.41	14.79	0.368
	(0.632)	(0.00708)	(0.00172)	(0.0803)	(12.44)	(12.43)	(0.296)
$\Delta C\times$ Post	-11.31***	-0.0113	0.00939	-1.758***	-108.4**	-97.05**	-2.852*
	(3.621)	(0.0414)	(0.00818)	(0.511)	(46.64)	(46.40)	(1.531)
$\Delta(Q \times C)\times$ Post	0.0630	0.000299	0.000493	0.0687**	0.0670	0.00406	-0.0845
	(0.349)	(0.00250)	(0.000737)	(0.0313)	(3.817)	(3.851)	(0.201)
R-Squared	0.415	0.509	0.505	0.547	0.453	0.451	0.454
Outcome Mean	679.26	6.32	2.2	142.78	5503.86	4824.6	4824.6
Average Impact	-16.34	.02	.04	-2.9	-159.34	-143	-5.08
Observations	41972	41972	41972	41972	41972	41972	41972

Standard errors in parentheses

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

Table 5: Effects of Quality and Cost Efficiency on Spending Outcomes

Notes: This table shows the impacts of quality and cost efficiency on spending outcomes. Column 1 reports the impacts on total spending: the amount paid for visits with the modal primary care physician by both the patient and the insurer. Column 2 reports impacts on the number of service received by the modal primary care physician. The outcome in column 3 is the average number of services received per visit, and column 4 reports the average price per service: total amount paid divided by total number of services. Columns 5 and 6 explore possible spillovers in spending on other providers. The outcome for the regression in column 5 is spending over all providers, not just the modal primary care provider. Column 6 narrows down to spending over other providers, not including the modal primary care provider. Column 7 displays impacts on the portion the patient paid out of pocket. The specification includes an interaction term between cost efficiency and quality scores to allow for complementarities.

	(1) Fractures	(2) Poisonings	(3) Childbirth
$\Delta Q\times$ Post	0.00138	-0.000172	0.00000737
	(0.00130)	(0.000141)	(0.00000776)
$\Delta C\times$ Post	0.00402	0.000421	0.0000211
	(0.00458)	(0.000500)	(0.0000243)
R-Squared	0.329	0.277	0.335
Outcome Mean	.0857	.00554	.00094
Average Impact	.01018	.0004	.00006
Observations	9150	9150	9150

Standard errors in parentheses

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

Table 6: Effects of Quality and Cost Efficiency on Placebo ED Outcomes

Notes: This table shows the impacts of quality and cost efficiency on each of the three components of “placebo” emergency department visits: fractures, poisonings, and childbirth. No statistically significant impacts are detected.

	(1)	(2)	(3)
	Preventable ED	Non-Emergency ED Visits	Placebo ED Visits
$\Delta Q\times$ Post	-0.000614	-0.000327	0.000328
	(0.000673)	(0.000262)	(0.000363)
$\Delta C\times$ Post	-0.000880	0.00145	0.000282
	(0.00743)	(0.00181)	(0.00167)
R-Squared	0.404	0.270	0.207
Outcome Mean	.09098	.02782	.02584
Observations			
N	14090	14090	14090

Standard errors in parentheses

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

Table 7: Effects of Quality and Cost Efficiency on ED Outcomes: Relative time balanced panel

Notes: This table shows the impacts of quality and cost efficiency on emergency department outcomes for the sample which is balanced on time relative to the switch (rather than calendar year)

	(1)	(2)	(3)
	Preventable ED	Non-Emergency ED Visits	Placebo ED Visits
$\Delta Q\times$ Post	-0.0000395	-0.0000552	0.0000505
	(0.000109)	(0.0000463)	(0.0000573)
$\Delta C\times$ Post	-0.000186	0.0000224	0.000298
	(0.000627)	(0.000277)	(0.000333)
R-Squared	0.345	0.210	0.156
Outcome Mean	.07384	.02414	.02222
Average Impact	-.00026	0	.00042
Observations	632450	632450	632450

Standard errors in parentheses

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

Table 8: Effects of Quality and Cost Efficiency on ED Outcomes: All Switches

Notes: This table shows the impacts of quality and cost efficiency on emergency department outcomes for the sample which includes endogenous switches.

	(1) Total Paid	(2) Total Paid	(3) Total Paid	(4) Total Paid	(5) Total Paid	(6) Total Paid
$\Delta Q\times$ Post	0.707	2.071**	2.775	0.157	0.461	-0.735
	(0.876)	(0.858)	(3.167)	(0.280)	(0.593)	(1.211)
$\Delta C\times$ Post	-13.20**	-8.129*	-30.13***	-25.66***	-15.78***	-14.47**
	(5.471)	(4.584)	(3.912)	(2.682)	(3.806)	(6.272)
Age Range	18-25	26-35	36-45	46-55	56-64	65 and older
Average Impact	-15.1	-7.64	-38.62	-35.84	-21.8	-21.24
R-Squared	0.370	0.418	0.298	0.393	0.264	0.373
Observations	17800	32907	98398	182127	204491	68411

Standard errors in parentheses

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

Table 9: Effects of Quality and Cost Efficiency over Age: All Switches

Notes: This table shows the impacts of quality and cost efficiency on spending outcomes for the full sample, which includes endogenous switches.

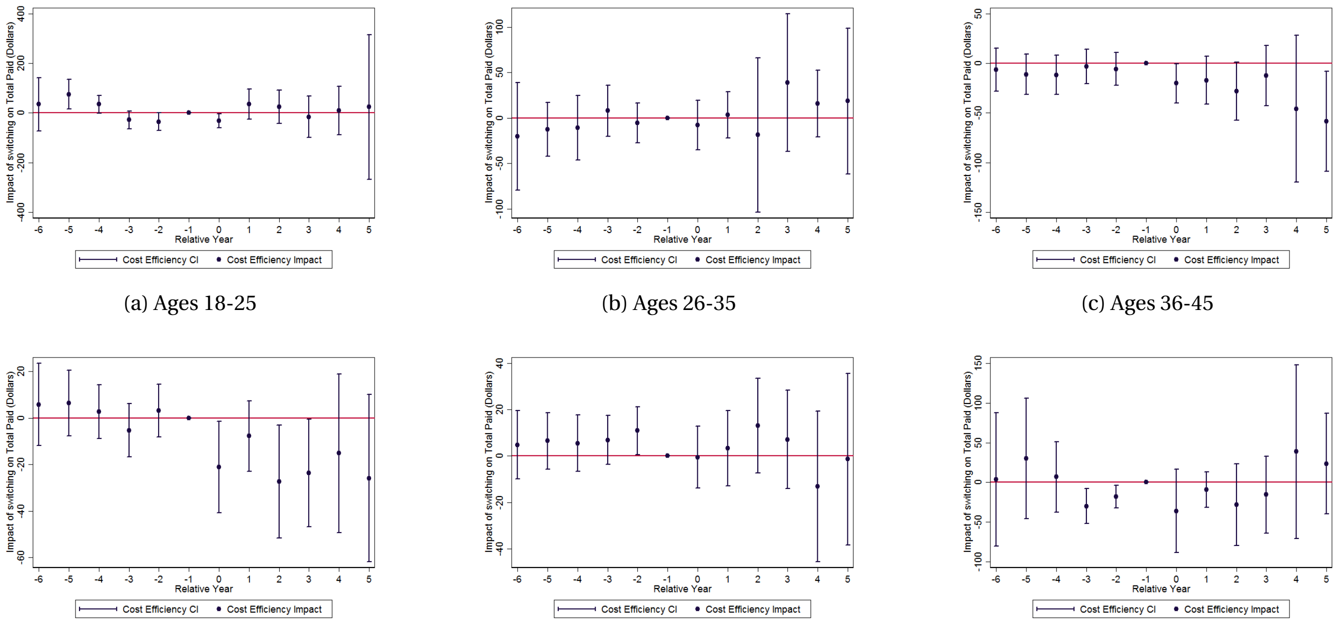
	(1) Total Paid	(2) Total Paid	(3) Total Paid	(4) Total Paid	(5) Total Paid	(6) Total Paid
$\Delta Q\times$ Post	8.609*	3.793*	4.376	-0.755	0.738	2.496
	(4.712)	(2.262)	(2.914)	(0.950)	(1.634)	(4.040)
$\Delta C\times$ Post	12.78	11.72	-34.20**	-11.00	-13.79	-19.23
	(18.00)	(16.56)	(15.74)	(11.19)	(9.002)	(20.59)
Age Range	18-25	26-35	36-45	46-55	56-65	65 and older
Average Impact	41.46	19.7	-43.02	-16.86	-18.1	-4.16
R-Squared	0.441	0.375	0.467	0.452	0.494	0.354
Observations	460	735	1755	3870	5075	1795

Standard errors in parentheses

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

Table 10: Effects of Quality and Cost Efficiency over Age: Relative Time Balanced

Notes: This table shows the impacts of quality and cost efficiency on spending outcomes for the sample which is balanced on time relative to switch. These results are generally under-powered; however, the pattern of results remains.



(a) Ages 18-25 (b) Ages 26-35 (c) Ages 36-45

(d) Ages 46-55 (e) Ages 56-64 (f) Ages 65 and up

Figure 7: Effect of quality and cost efficiency on total amount paid by age group

Notes: This figure displays the impact of cost efficiency on total amount paid separately by age groups.

	(1)	(2)	(3)	(4)	(5)	(6)
	Preventable ED	Preventable ED	Preventable ED	Preventable ED	Preventable ED	Preventable ED
$\Delta Q\times$ Post	0.000215	-0.0000366	0.0000654	-0.000268	-0.00000499	0.000254
	(0.000222)	(0.000203)	(0.000200)	(0.000203)	(0.000243)	(0.000399)
$\Delta C\times$ Post	0.00576*	-0.00317	0.00134	0.000791	-0.00202*	0.000374
	(0.00344)	(0.00259)	(0.00133)	(0.000966)	(0.00121)	(0.00272)
Age Range	18-25	26-35	36-45	46-55	56-64	65 and older
R-Squared	0.379	0.411	0.327	0.338	0.323	0.393
Observations	17800	32907	98398	182127	204491	68411

Standard errors in parentheses

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

Table 11: Effects of Quality and Cost Efficiency over Age: All Switches

Notes: This table shows the impacts of quality and cost efficiency on preventable emergency room visits for the full sample, which includes endogenous switches.

	(1)	(2)	(3)	(4)	(5)	(6)
	Preventable ED	Preventable ED	Preventable ED	Preventable ED	Preventable ED	Preventable ED
$\Delta Q\times$ Post	-0.00542	0.000884	-0.000618	-0.000927**	0.000160	-0.00172
	(0.00684)	(0.00691)	(0.00101)	(0.000466)	(0.00169)	(0.00192)
$\Delta C\times$ Post	0.136	-0.0843	-0.00685	0.00289	-0.0113**	0.0203
	(0.130)	(0.0544)	(0.00445)	(0.00519)	(0.00510)	(0.0359)
Age Range	18-25	26-35	36-45	46-55	56-64	65 and older
R-Squared	0.395	0.491	0.386	0.379	0.360	0.482
Observations	460	735	1755	3870	5075	1795

Standard errors in parentheses

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

Table 12: Effects of Quality and Cost Efficiency over Age: Relative Time Balanced

Notes: This table shows the impacts of quality and cost efficiency on preventable emergency room visits for the sample which is balanced on time relative to switch.

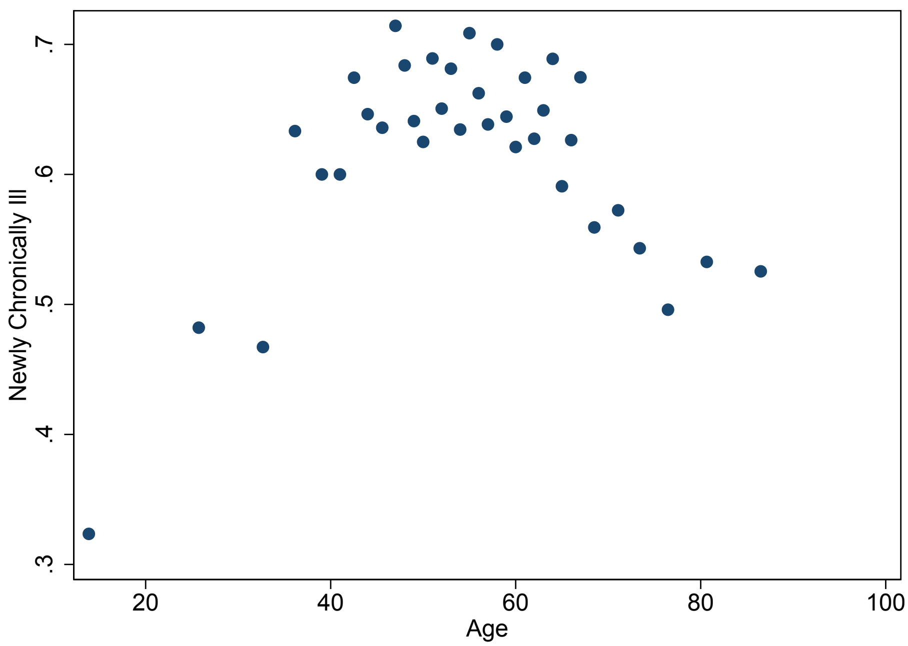
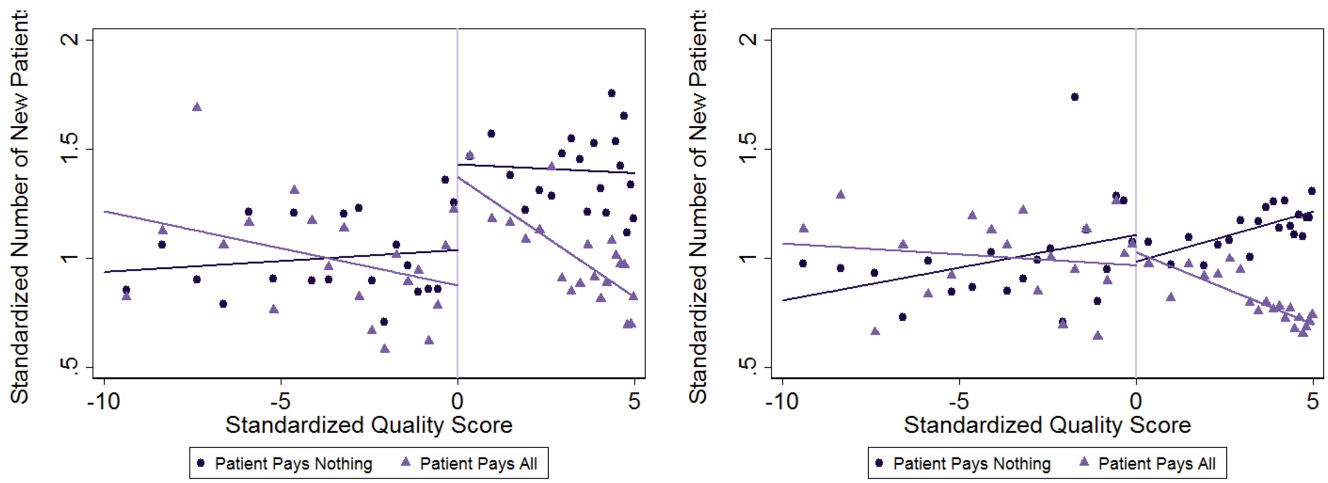


Figure 8: Binned Scatterplot: Newly Chronically Ill vs Age

Notes: This figure displays a binned scatterplot of the proportion of patients who became chronically ill during the timeframe studies against teh age of patients. The patients who were most likely to become chronically ill were middle-aged, between about 40 and 70 years of age.

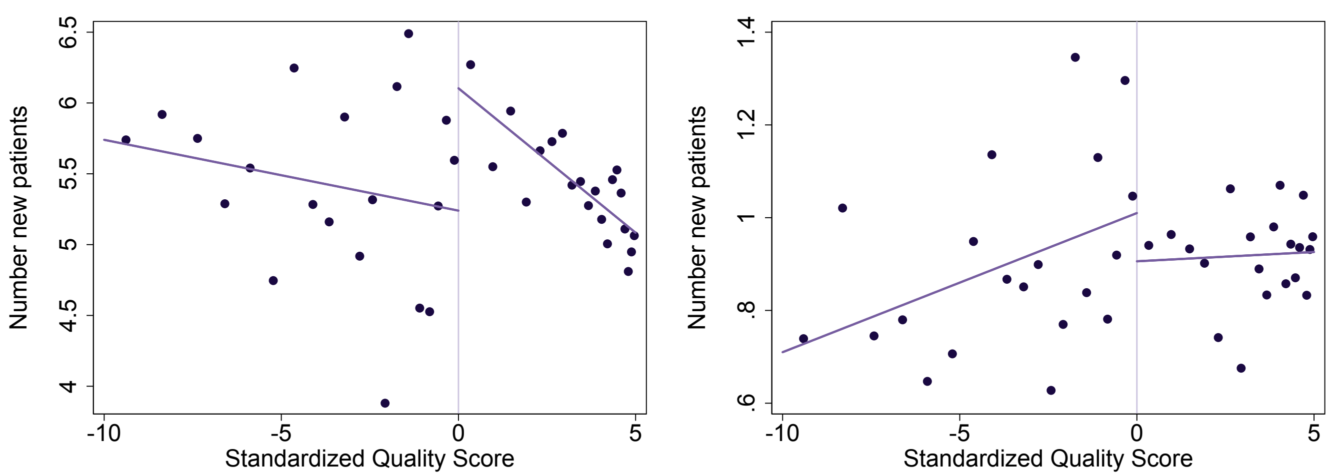


(a) Zero vs. Two Hearts

(b) Zero vs. One Heart

Figure 9: Effects by Proportion Paid by the Patient

Notes: This figure displays impacts of the Premium and Quality designations broken down by the amount patients paid, for patients who paid nothing out-of-pocket (Patient Pays Nothing) compared to patients who covered 100% of their care (Patient Pays All).



(a) Not at Capacity

(b) At Capacity

Figure 10: Effects by Provider Capacity

Notes: This figure displays impacts of the Premium designation separately for physicians who are not at capacity in Panel A versus those who are at capacity in Panel B. Capacity is proxied as follows: A physician is considered to be at capacity if they did not see any new patients but saw a positive number of returning patients during the time period just before new designations were released.

	(1)	(2)	(3)	(4)
	Total Paid	Total Paid	Total Paid	Total Paid
$\Delta Q\times$ Post	1.664***	1.777***	1.710***	1.803***
	(0.547)	(0.549)	(0.544)	(0.547)
$\Delta C\times$ Post	-11.18***	-11.22***	-11.28***	-11.17***
	(3.670)	(3.664)	(3.657)	(3.653)
relative switch year =-4	-17.06		-16.14*	
	(19.29)		(8.887)	
relative switch year =-3	-11.22		-24.01**	
	(32.41)		(10.19)	
relative switch year =-2	-5.682		-29.91**	
	(46.48)		(11.87)	
relative switch year =-1	-86.95		-129.5***	
	(60.03)		(11.82)	
relative switch year =0	40.73		-17.49	
	(73.56)		(13.95)	
relative switch year =1	38.48		-29.99	
	(89.61)		(19.26)	
relative switch year =2	46.74		-36.40*	
	(102.5)		(19.44)	
relative switch year =3	63.10		-31.56	
	(117.2)		(22.31)	
relative switch year =4	88.19		-27.06	
	(133.1)		(30.28)	
relative switch year =5	144.3		11.85	
	(147.2)		(31.69)	
post-switch		95.82***		29.77***
		(13.91)		(9.742)
Constant	675.3***	642.6***	715.2***	713.8***
	(53.71)	(5.147)	(0.267)	(0.109)
R-Squared	0.415	0.414	0.596	0.596
Outcome Mean	679.26	679.26	714.16	714.16
Sample	Induced Switches Only	Induced Switches Only	All Switches	All Switches
Observations	41972	41972	1385461	1385461

Standard errors in parentheses

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

Table 13: Effects of Quality and Cost Efficiency on Spending

Notes: This table shows the impacts of quality and cost efficiency on total spending. Column 1 displays the coefficients on relative switch year indicators, while column 2 displays the impacts from a specification which collapses relative year indicators into a single post-switch indicator. Columns 3 and 4 repeat these analyses on the full sample which includes endogenous switches.

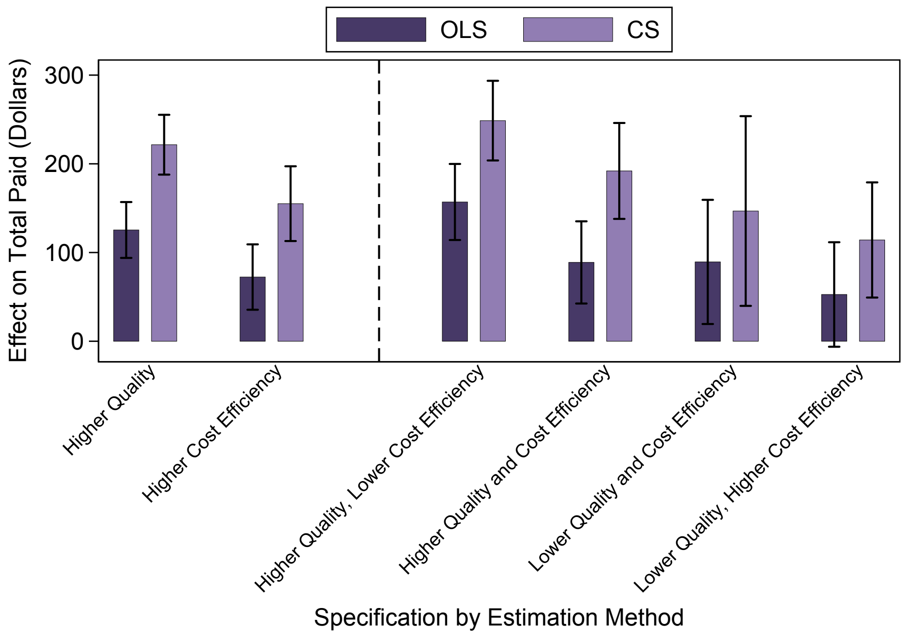


Figure 11: Comparison of Effect of Switching within Sub-Groups: OLS vs CS

Notes: This figure plots impacts of switching over sub-groups, exploring the impacts of switching separately for patients whose switch was to a higher-quality provider, then for patients whose switch was to a more cost-efficient provider. The next four sets of estimates subset further, first showing the impact of switching for those who switch to higher quality, less cost efficient providers, then those who switch to higher quality and cost efficiency, lower quality and cost efficiency, and finally lower quality but higher cost efficiency. While Callaway-Sant'Anna (CS) estimates are generally larger than OLS estimates, both sets of results point to the same takaway, that switching to higher quality, lower cost efficiency providers is more expensive than switching to lower quality, more cost efficient providers.