NBER WORKING PAPER SERIES

# PHYSICIAN GROUP INFLUENCES ON TREATMENT INTENSITY AND HEALTH: EVIDENCE FROM PHYSICIAN SWITCHERS

Joseph J. Doyle Jr.
Becky Staiger

Working Paper 29613
http://www.nber.org/papers/w29613

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2021

We thank Leila Agha, Laurence Baker, David Chan, David Cutler, Phillip Decicca, William Evans, Daniel Hungerman, Anupam Jena, Ethan Lieber, David Molitor, Mohan Ramanujan, Jonathan Skinner, Tara Templin and seminar participants at ASHEcon and Notre Dame for helpful comments and suggestions. We gratefully acknowledge support from the National Institutes of Health R01 AG41794. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Physician Group Influences on Treatment Intensity and Health: Evidence from Physician Switchers
Joseph J. Doyle Jr. and Becky Staiger
NBER Working Paper No. 29613
December 2021
JEL No. I10

## ABSTRACT

Treatment intensity varies remarkably across physicians, yet the key drivers are not well understood. Meanwhile, the organization of healthcare is undergoing a secular transformation as physicians increasingly work in groups. This paper tests whether physicians' group affiliation matters for practice styles and patient health. Using Medicare inpatient claims data, we compare these outcomes before and after physicians switch between groups of varying treatment intensity while remaining in the same hospital to control for practice setting. Event studies show that internists who join more-intensive groups immediately increase their own treatment intensity, with an elasticity of approximately 0.3; the opposite is found for internists who switch to groups that are less intensive. This change in Medicare spending largely stems from greater quantities of care provided, with some evidence of a change in coding behavior. We do not detect a change in health outcomes, suggesting that treatment intensity induced by group affiliation may not be productive.

Joseph J. Doyle Jr.
MIT Sloan School of Management
100 Main Street, E62-516
Cambridge, MA 02142
and NBER
jjdoyle@mit.edu

Becky Staiger
Stanford School of Medicine
Center for Biomedical Informatics Research
1265 Welch Rd
Stanford, CA 94305
bstaiger@stanford.edu

A data appendix is available at http://www.nber.org/data-appendix/w29613

# 1 Introduction

It is well known that physicians vary remarkably in their treatment intensity across regions, and even within the same hospital (Fisher et al., 2003a; Tsugawa et al., 2017). While addressing the drivers of this wide variation could improve productivity and slow the growth of healthcare spending, they are not well understood (Berndt et al., 2015; Chan, 2021; Cutler et al., 2019; Epstein and Nicholson, 2009). Recent research has started to fill in this gap, identifying potential sources of variation that include demand-side factors, such as patient illness severity and preferences; and supply-side factors, such as physician practice styles and institutional constraints and incentives (Finkelstein et al., 2016; Cutler et al., 2019).

Meanwhile, the organization of healthcare is undergoing a long-term transformation as physicians increasingly work in groups (Capps et al., 2017; Kane, 2017). By 2016, we find that 84% of physicians in the Medicare system work in groups, including 93% of inpatient physicians, and these shares have been growing over time. This secular trend away from solo practice is explained by a variety of factors that include growing administrative burdens and incentives to coordinate care. Despite considerable discussion of this supply-side shift in healthcare organization, the implications for treatment intensity and patient health merit more attention (Heeringa et al., 2020; Zwiep et al., 2021; Muhlestein and Smith, 2016; Welch et al., 2013). Moreover, a better understanding of group influence on physician practice style would not only inform the implications of the trend towards group practice, but may also help shed light on a fundamental question in healthcare productivity: what are the key drivers of practice-style variation across physicians?

This paper tests whether a physician's group affiliation influences treatment intensity and health outcomes. Comparing physicians who belong to groups that vary in their practice styles could be misleading due to endogeneity bias: patients treated by different groups may vary in their illness severity or preferences for treatment; physicians with similar practice styles may choose to form a group, leading to reverse causality; and physicians may influence one another simultaneously, leading to a reflection problem (Manski, 1993). Our empirical strategy aims to circumvent these endogeneity concerns in three ways. First, using event studies, we compare physicians' treatment intensity before and after they switch group practices. Some physicians choose to switch to practices that are more intensive than their origin group, while others choose to switch to less-intensive groups. An abrupt and stable change in treatment intensity upon switching groups suggests that either the group exerts some influence on treatment intensity directly, or a physician may need to switch groups in order to practice in a different way. Either way, such a change points to treatment variation stemming in part from group affiliation. Second, our main results focus on physicians practicing internal medicine to avoid comparisons across different types of specialists. Third, comparisons are made among those who switch groups but remain in the same hospital. By using within-hospital variation, the analysis holds constant characteristics about the facility that are unchanging over time.

To make these comparisons, we investigate Medicare inpatient claims detailing the treatment of beneficiaries from 2008 to 2016. We characterize physician treatment intensity, as well as the intensity of other physicians in their group, using average Medicare payments each quarter. We find that a physician who joins a more-intensive group immediately increases her own intensity at the time of the switch. In particular, joining a group that has a one-standard deviation higher group intensity (approximately 78 log points, or an increase in per-physician reimbursement of approximately 120%), results in a 23 log-point increase in physician intensity. We find the estimates are symmetric for moves to relatively higher- or lower-intensity groups. Using a more parsimonious model, our results imply that approximately 18% of the within-hospital variation in observed intensity across physician groups can be attributed to group-specific factors, while the remaining 82% is attributable to physician-side components.

To explore how patient welfare may be impacted by such group effects, we also evaluate how this change affects several quality-of-care measures, including hospital readmissions and mortality. Despite a change in treatment intensity that scales with the change in the group-intensity measure, we do not detect a change in these health outcomes. While other aspects of care can vary with group intensity, these results suggest that the group-induced increases in intensity may not be productive.

A key identifying assumption when employing an empirical strategy that studies physicians who switch groups is that they do not experience contemporaneous shocks that are correlated with the size of the change in group intensity. For example, patient characteristics could differ across origin and destination groups, or physicians may change their preferred practice style precisely at the time when they switch practices. We find that observable patient characteristics are similar before and after the moves, which is consistent with within-hospital changes in group affiliation not affecting the types of patients that physicians treat. In addition, we observe that the trajectory of treatment intensity prior to the switch is unrelated to the change in group intensity, which provides additional confidence that the identifying assumption is plausible. Note that we focus on switching physicians and rely on the gradient of the change in group intensity to identify the effects on physician intensity. As a result, switching physicians are allowed to differ from those who do not switch, and we describe the differences across these physicians to begin to learn whether the findings for switchers are likely to apply to non-switchers.

We explore for heterogeneity of effects along two dimensions: across different types of physicians and across different types of groups. In terms of physicians, we find that the effects are particularly large for internists, though still present among a non-internist sample. We find similar results at different points in physicians' careers, suggesting that the results are not driven by incentives to switch groups at particular times, such as pre-retirement moves. We also observe relatively similar results across different levels of pre-switch physician intensity, suggesting that the influence of a group environment on physicians is not dependent on a physician's own practice style.

Next, we recognize that other features of the groups besides intensity vary when a physician makes a switch. We use our empirical strategy to compare physicians who join groups that vary along another dimension that we can measure: group size. Here, we find suggestive evidence that switching to a larger group is associated with a reduction in healthcare spending. We do not find differences in health outcomes from joining a larger group, although these estimates are not as precisely estimated. This combination of lower spending and similar health outcomes is reassuring given policy concerns that group consolidation may lead to greater healthcare spending. Meanwhile, our main results are robust to a number of checks, including when we compare physicians who join more- or less-intensive groups across a wide range of destination-group sizes.

Last, we explore potential mechanisms for the group's influence. Physicians appear to provide a higher quantity of care to their patients in terms of a larger number of claims overall and per-patient, as well as claims that are coded to indicate greater intensity. In particular, the most common types of claims for internists are "evaluation and management" claims (Cabral et al., 2021). We find that more-intensive groups have such claims that are coded to represent a greater amount of time and effort spent treating the patients, and switching to a more-intensive group leads to a sudden and sustained increase in such coding. This could reflect greater intensity, a change in coding, or both. Meanwhile, average predicted one-year mortality for the patients the physician treats does not change across the group switch, suggesting that this increased intensity is not reflective of a change in patient need. Coupled with similar health outcomes despite higher levels of billing, the results suggest that group-induced increases Medicare spending may not be productive. Overall, we find that while most of the treatment variation stems from physician-specific factors, group affiliation is an important component of the variation in treatment intensity among inpatient physicians.

The rest of the paper proceeds as follows. Section 2 briefly describes the trend in practicing in groups over time and related literature. Section 3 introduces our empirical framework. Section 4 describes our data and details the sample construction. Section 5 presents the results and Section 6 discusses potential mechanisms. Section 7 concludes.

# 2 Related Literature

## 2.1 Physician Group Formation

There are growing incentives for physicians to practice in groups. Market and environmental factors—such as increasing financial burdens associated with medical debt, administrative requirements including quality reporting and documenting meaningful-use of health information technology, and policies that generate new incentive structures for more coordinated care—all have prompted a significant shift toward group practices (Harris, 2010; Kane, 2017; Muhlestein and Smith, 2016; Welch et al., 2013). Indeed, most physicians now work in groups. In Medicare claims data that we describe in more detail below, we see a steady decline in the share of physicians practicing in solo practice over time.[1] Figure A.1 shows that by 2016, 7% of physicians in the inpatient setting were working in a solo practice, down from 15% in 2008.

Several mechanisms have been proposed to explain how groups can improve performance, including that groups can (1) benefit from economies of scale, such as the incorporation of health IT, and (2) alter compensation models to reward quality of care in addition to the quantity of care provided. As noted in the introduction, there is extensive discussion gauging the influence of group affiliation on physicians, but effects on treatment decisions and patient health have been lagging (for a review, see Zwiep et al. (2021)). Epstein et al. (2010) show that compared to solo practitioners, group practices in obstetrics are better able to match patients to specialists, improving their health. Similarly, for inpatient cardiac care, Ketcham et al. (2007) find that patients treated by physicians in solo practices are less likely to receive invasive procedures and have higher mortality.

Much of the existing research of physician groups focuses on group size. Spending and quality measures have been compared across different-sized groups controlling for practice and physician char-

[1] As detailed below, we define a group as those who bill using the same (de-identifed) tax identification number (TIN). Our billing data come from a 20% random sample of Medicare beneficiaries. As a result, the share in solo practice is likely even higher, although the trend toward fewer solo practices is evident in the 20% sample

acteristics using a selection-on-observables approach. For example, Casalino et al. (2014) combine survey data on group-practice size with Medicare quality measures and find that small practices have 30% lower preventable-admissions compared to practices with 20 or more physicians. McWilliams et al. (2013) provide more nuanced evidence of larger hospital-based practices providing greater treatment intensity and higher readmission rates, while larger independent physician groups have higher quality scores and lower spending levels. Such comparisons could reflect differences in patient characteristics as group types and sizes vary.

In general, the ongoing wave of physician-group formation and consolidation is striking. Even within the hospital setting, Figure A.1 shows that average group size in our sample in the first quarter of 2008 was 67 and grew nearly three-fold to 184 by the end of 2016. This growth in size motivates our exploration of robustness of the main results to changes in group size, as well as a direct examination of the effects of group size on treatment and patient health.

While the existing literature examining the relationship between group structure and physician performance has primarily focused on group size, our study adds to the discussion by considering another element of the group environment: group-practice intensity. We focus our exploration on the inpatient setting in order to consider short-term, welfare-relevant outcomes including spending, readmissions, and mortality. Exploring group switches within a hospital also allows us to control for fixed attributes of the hospitals where physicians practice.

## 2.2 Variation in Treatment Intensity

The ongoing trend and ubiquity of group practice raises the question of whether a physician's group matters for the way they treat patients. The variation in treatment intensity across physicians is remarkable, even among physicians working in similar practice environments and treating similar patients. Wennberg and Gittelsohn (1973) famously showed that tonsillectomy rates varied widely across Vermont towns, launching a large literature documenting remarkable small-area variation in treatment intensity. Potential drivers of this variation include the preferences and training of physicians (Cutler et al., 2019; Epstein and Nicholson, 2009), along with institutional features such as financial incentives, constraints, and practice norms (Clemens and Gottlieb, 2014; Molitor, 2018). Tsugawa et al. (2017) demonstrate that among general internists treating Medicare patients within the same hospitals, physicians at the 90th percentile of spending had 50% higher hospitalization costs compared to the 10th percentile, even after adjusting for patient characteristics. They show that this is relatively larger than the substantial between-hospital, cross-region variation in treatment intensity (Baker et al., 2014a; Barnato et al., 2007; Cutler et al., 2019; Finkelstein et al., 2016; Fisher et al., 2003a,b).

Cutler et al. (2019) explore the black box of "supply-side" drivers of regional variation by using physicians' answers to vignettes of patient cases to identify factors that influence physician behavior. They find that approximately 60% of the variation in end-of-life spending across markets can be explained by whether a physician is classified as a "cowboy" (more aggressive) or a "comforter" (less aggressive), and that physician beliefs regarding the efficacy of therapeutic interventions (not necessarily based on clinical effectiveness) are the key drivers of these differences in intensity, explaining as much as 35% of end-of-life expenditures. While the authors find that group structure (namely, size and single- or multi-specialty practice) explains only a small amount of the variation in physician behaviors, our analysis extends this analysis of group effects by explicitly examining the intensity of the group, which is not captured in the surveys, and may be more relevant to a physician's own intensity.

One source of influence on practice styles may be the physician's peers. Epstein and Nicholson (2009) study how residency training and a physician's local peers (in the same hospital or in the same market) might affect a physician's propensity to opt for a C-section (instead of vaginal birth) during delivery. The authors find only a very small effect of both training and local C-section rates on a physician's own C-section rate, where residency programs explain approximately 2% of variation. They conclude that much of the practice variation between physicians is likely due to a physician's beliefs regarding the efficacy or appropriateness of specific treatments. They also document a significant amount of within-region variation in C-section rates, observing that within-market variation is approximately twice as large as variation between markets, although the implications for patient welfare are unclear.[2] Saghafian et al. (2019) and Chan (2016) study emergency-room physicians who practice side by side, finding that physicians who work with faster or higher quality peers tend to perform worse, while holding physicians jointly responsible for their care can reduce a "foot-dragging" form of moral hazard when patients are assigned independently across physicians.

Group-practice affiliation can affect physician financial incentives via different compensation models, and there is evidence that physicians respond to such financial incentives (McGuire and Pauly, 1991). For example, Clemens and Gottlieb (2014) exploit a regional consolidation of Medicare fees that resulted in significant changes in payment rates across areas to explore the role of financial incentives in physician

[2]If there is an "optimal" level of C-section frequency, then within-market variation in rates implies that some patients will receive more/less C-sections than recommended, thereby reducing patient welfare. If, however, variation reflects differences in patient preferences or suitability for C-section, then variation could be welfare enhancing.

treatment decisions. They find that a 2% increase in payment rates led to a 3% increase the provision of care, and that the use of elective procedures was more responsive to this change than non-elective procedures. Alexander (2015) and Alexander and Schnell (2019) use fee changes in Medicaid to show that physicians respond to increased payments by increasing their use of C-sections and increasing the number of Medicaid patients they treat, respectively. Cabral et al. (2021) find similar elasticity estimates when examining a payment reform that increased the generosity of Medicaid payments for "evaluation and management" visits.

## 2.3 Movers Analyses

Closely-related work studies regional variation in treatment and outcomes. Molitor (2018) studies cardiologists who move to a more (or less) intensive area and tests whether this results in a change in treatment intensity. He finds that the environment in which a cardiologist practices (which describes factors such as hospital capacity and productivity spillovers) accounts for 60-80% of the observed variation across hospital referral regions (HRRs) in catheterization rates. Notably, he observes that the effect of a change in intensity in more localized environments (i.e. hospitals) on a physician's own catheterization rate is larger than the effect of a change in intensity at the broader geographic region, suggesting that physician behavior may be especially sensitive to small-area environments. Similarly, Finkelstein et al. (2016) study patient movers among Medicare beneficiaries to decompose regional variation in utilization into demand-side and place-specific, supply-side factors. They find that patient-specific components (such as health and preferences) account for approximately 40-50% of observed variation in healthcare utilization across HRRs.

While these papers focus on cross-region variation, we use similar methods to explain within-hospital variation in treatment intensity, holding constant other variation-contributing factors that might otherwise change when individuals move across regions, such as area-level resources and system-level preferences for care intensity.[3] The methodology also has the advantage of providing a straightforward decomposition of variation into group-specific and physician-specific factors, as detailed in the empirical framework.

[3] Agha et al. (2020) note that identifying effects of a healthcare environment using an agent's move across regions can be challenging in terms of differentiating the effect of an environmental change from the effect of other elements of the regions that are correlated with the environment being studied.

# 3 Empirical Framework

## 3.1 Estimating Group Effects on Physician Behavior

Our goal is to test whether a physician's group matters for how they practice using various measures of treatment intensity. Appendix B includes a simple model of intensity choice that is the result of physician and group effects, taking into account patient characteristics as well. Physicians influence treatment intensity due to their preferences, skill, private (opportunity) costs of administering the care, and their beliefs about the effectiveness of the care (e.g. Ellis and McGuire (1986); Alexander (2015); Clemens and Gottlieb (2014); Cutler et al. (2019)). Groups can influence treatment decisions through productivity incentive structures, billing technology, and the group's relative weighting on profits versus benefits to patients (Dafny, 2005; Song et al., 2020). The end result is a straightforward model of physician intensity in the spirit of Abowd et al. (1999) (hereafter, referred to as "AKM") and Finkelstein et al. (2016) that includes physician (worker) and group (firm) fixed effects.

Abstracting from time-varying characteristics of the environment and patients, the following simplified model of a physician's observed level of intensity in terms of these effects for physician $p$ and group $g$ can be written as:

$$y_{pg} = \alpha_p + \delta_g + \varepsilon_{pg} \tag{1}$$

Where $\alpha_p$ are physician fixed effects; $\delta_g$ are group fixed effects; and $\varepsilon_{pg}$ are unobserved characteristics that drive variation in intensity, such as patient characteristics. For those physicians who switch to group $g'$ then:

$$y_{pg'} = \alpha_p + \delta_{g'} + \varepsilon_{pg'}$$

Our empirical approach compares physicians before and after they switch groups to physicians who do not switch groups. The main idea is that physician effects are constant across the switch to a new group. If moves are exogenous, allowing us to ignore the $\varepsilon$ terms, then the change in observed intensity will identify the difference of group effects, as the physician fixed effects are assumed to be unchanging:

$$E(y_{pg'} - y_{pg}) = \delta_{g'} - \delta_g$$

To estimate these group effects, we relate the change in treatment intensity of switching physicians when they join more- or less-intensive groups. To characterize this change in group intensity, we calculate the degree to which the intensity of the destination group differs from the origin group as:

$$\Delta_{pmt(p)} = \overline{y}_{d(p),q<0} - \overline{y}_{o(p),q<0} \tag{2}$$

where $\overline{y}_{o(p),q<0}$ and $\overline{y}_{d(p),q<0}$ are the average Medicare reimbursement of the other physicians in the origin and destination groups, respectively, calculated in the four quarters, $q$, prior to the switch, where the quarter of the switch is normalized to zero. In terms of notation, $\Delta_{pmt(p)}$ is defined specifically for each switching physician; however, in our models, we omit the $p$ in the subscript for simplicity's sake. Given that a physician can impact her peers in the origin group, we also estimate models that relate the change in physician behavior solely to the destination group intensity measured prior to the physician's arrival. In addition, we will examine how these groups differ along other dimensions, such as group size.

Note that from Equation 1 the change in group environment represents a change in average physician effects (from other physicians) and group effects as we average over physicians to construct $\Delta_{pmt}$:

$$\Delta_{pmt} = \overline{\alpha}_{d(-p),q<0} - \overline{\alpha}_{o(-p),q<0} + \delta_{d(p),q<0} - \delta_{o(p),q0} \tag{3}$$

To define the share of the variation in group intensity that stems from physician effects vs. group effects, simply divide both sides by $\Delta_{pmt}$:

$$\begin{aligned} Share_g =& (\delta_{d(p),q<0} - \delta_{o(p),q<0})/\Delta_{pmt} \\ Share_p =& (\overline{\alpha}_{d(-p),q<0} - \overline{\alpha}_{o(-p),q<0})/\Delta_{pmt} \end{aligned} \tag{4}$$

As in Finkelstein et al. (2016), we use an event-study approach in which physician $p$ switches from origin group $o$ to destination group $d$ to recover the average effect of group intensity on physician's own intensity. We discuss any contamination of the estimates that might arise from using staggered treatment timing below. In this empirical strategy, the jump in a physician's intensity at the time of the switch identifies the extent of the influence of the group environment on a physician's own intensity. Using the definitions in Equation 4 and the timing of the switch, the AKM model can be re-written for event time $r$ as:

$$\begin{aligned} y_{pg} =& \alpha_p + \delta_{d(p),q<0} + \mathbb{1}(r > 0)(\delta_{d(p)} - \delta_{o(p)}) + \varepsilon_{pg} \\ =& \alpha_p + \delta_{d(p),q<0} + \mathbb{1}(r > 0) Share_g * \Delta_{pmt} + \varepsilon_{pg} \end{aligned} \tag{5}$$

where $\mathbb{1}(r > 0)$ is an indicator for the post-switch period.

When we bring this model to the data, we include controls for time varying characteristics of the environment and patients, and our estimating sample includes non-switchers to help estimate the relationships associated with these controls. Our estimating equation models physician $p$'s treatment intensity (or quality of care), $y$, in group $g$ and hospital $h$ during calendar-quarter $t$ as:

$$y_{pght} = \alpha_p + \beta_{ht} + \sum_{q=-10}^{10} \gamma_q \mathbb{1}\{Q_{pt} = q\} + \sum_{q=-10}^{10} \theta_q \mathbb{1}\{Q_{pt} = q\} \times \widetilde{\Delta}_{pmt} + \lambda X_{pt} + \varepsilon_{pght} \tag{6}$$

where $\alpha_p$ is a physician-episode fixed effect (where "episode" refers to the period of at least nine quarters before and after a switch to a new group at $q = 0$) to control for time invariant characteristics of the physician, including the types of groups that are chosen by physicians and their tenure in the data when we move to longer time horizons and the panel becomes unbalanced.[4] $\beta_{ht}$ represents hospital-specific calendar year-quarter fixed effects to control for hospital-specific trends that occur contemporaneously with or around the physician's switch. $X_{pt}$ are average patient characteristics measured at the physician-quarter level; and $\varepsilon_{pght}$ is an error term that we assume to be mean zero and mean independent of the event-time indicators, their interaction with relative group intensity, and included patient characteristics.

The remaining elements of the empirical model trace the outcomes of interest in the quarters to and from a switch. We are specifically interested in how the difference in intensity across the destination and origin groups, $\Delta_{pmt}$, affects the physician's behavior around the switch. Thus, the main coefficients of interest are the $\theta_q$s.

As is typical in interaction models, we de-mean $\Delta_{pmt}$, represented as $\widetilde{\Delta}_{pmt}$ to ease interpretation of the estimates. In practice, $\Delta_{pmt}$ has a mean that is close to zero. We set all indicator variables for the quarter relative to switch to 0 for the non-switching cohort.

The post-switch slope created by the $\theta_q$s in $q > 0$ is also informative. As Molitor (2018) discusses in his interpretation of post-move trends in catheterization rates, an immediate jump in $\theta_q$ followed by a relatively flat slope in the estimates of $\theta_q$s for $q > 0$ is consistent with group norms and policies driving the group influence. Conversely, an increasing slope (following an immediate jump at $q = 0$) may reflect a more long-term, adaptive group effect, suggesting that in addition to an immediate alteration of behavior due to environmental change, the physician continues to alter their behavior through learning

[4] Molitor (2018) uses HRR-level fixed effects instead of physician-level fixed effects in order to test for selection among moving physicians, namely that they are systematically different from baseline migrants in the same HRR. We will explore how those who switch groups differ from those who do not switch below.

over time or slower-acting peer influence.

The pattern we find in the event studies suggests that a more-parsimonious model that improves precision is also informative. That is, we estimate a pre-post version of the event study in which we define an indicator, Post Switch, as equal to one for all quarters $q \in [1, 10]$ and zero otherwise, in addition to an indicator, Qtr=0, which is equal to one when $q = 0$ and 0 otherwise. We include Qtr=0 to explicitly allow for a transition period during which a physician is practicing for part of the quarter in the origin group, with the remaining time in the destination group. This is motivated by a plot of the share of HCPCS associated with a given group in a particular quarter shown in Figure A.2. This specification tests for the effect of a change in group intensity relative to the pre-period, $q \in [-10, -1]$:

$$\begin{aligned} y_{pght} =& \alpha_p + \beta_{ht} + \delta_1 \mathbb{1}\{\text{Post Switch}\} \times \widetilde{\Delta}_{pmt} + \delta_2 \mathbb{1}\{\text{Qtr=0}\} \times \widetilde{\Delta}_{pmt} \\ &+ \delta_3 \mathbb{1}\{\text{Post Switch}\} + \delta_4 \mathbb{1}\{\text{Qtr=0}\} + \lambda X_{pt} + \varepsilon_{pght} \end{aligned} \tag{7}$$

Flat pre-switch trends estimated from Equation 6 would provide some evidence that our key identifying assumption of parallel, counterfactual trends in the post-switch period is plausible.

Using the above framework, we can also investigate changes in other group characteristics. One (particularly policy relevant) dimension is group size. We evaluate the effect of group size on physician treatment intensity in two main ways. First, we test whether our main results change when we control for the change in group size. Specifically, we estimate Equation 6 by quartiles of $\Delta_{size}$, which is the change in the number of physicians across the origin and destination groups. Second, we conduct a similar analysis as in our main specification, but we directly test whether the change in group size matters for treatment intensity. That is, we substitute the log size of the group for group intensity, such that $\Delta_{size}$ replaces $\Delta_{pmt}$ as the primary environmental change of interest.

## 3.2 Inference

We compute two-way clustered standard errors at the physician and group levels to incorporate correlation within and between these two attributes. Because $\Delta_{pmt}$ is a generated regressor, we also report similar standard errors when using a bootstrap procedure that incorporates the variability due to the calculation of $\Delta_{pmt}$. In addition, we report robustness to different minimum-observation requirements imposed to calculate group- and physician-intensity measures.

### 3.3 Threats to Identification

A key identification assumption of any difference-in-differences analysis is parallel trends—in our case, that absent a switch in groups, trends in $y$ among switching physicians in the period after the switch would be similar regardless of the magnitude of the change in group treatment-intensity. A feature of the event study is that we can evaluate whether physicians who join groups with different levels of intensity are on similar trends prior to the switch. Our setting features staggered episodes, which allows us to control for calendar-time effects. Sun and Abraham (2020) note that in such settings, estimates may not be a straightforward average across individuals if there is anticipation or heterogeneous treatment effects. We also show that the event study estimates using the pooled, staggered events provide a good summary of the set of event studies that are estimated separately for episodes defined by switches that happen in the same calendar quarter.

What the identifying assumption does not allow for is a shock to the physician's preferred level of treatment that coincides with the time of the move and is related to the size of the change in intensity across the groups. For example, characteristics of the physician's environment could change at the same time, such as patient characteristics or complementary labor and capital. In an attempt to control for such changes, we restrict our analysis to physicians who remain in the same hospital before and after switching groups. Even restricting our analysis to within-hospital group switches, the physician may begin treating different types of patients (Chang and Obermeyer, 2020). To investigate this concern, we test for balance of patient characteristics before and after the physicians switch groups. Specifically, we estimate Equation 6, replacing measures of intensity and quality of care on the left-hand side with several key patient characteristics that have been linked to differing levels of treatment intensity and health outcomes: average age, race, and sex. We also evaluate balance of predicted one-year mortality, which is based on these demographics and prior-year comorbidities, to investigate changes in patient complexity that might justify a change in treatment intensity.

A related concern is that a physician may choose to move to a group with higher (or lower) intensity in order to change their treatment intensity, such as physicians starting to taper their practice in preparation for retirement. Here, we again evaluate any presence of pre-trends to help us investigate whether behavior changes in anticipation of the switch. We also note that a sudden change in treatment style at the time of the switch suggests that physicians are constrained in their behavior until the move occurs, which would suggest that group affiliation matters for treatment intensity even if physicians choose a group because it is a better match for their preferred intensity level. For retirement influences in particular, we directly estimate effects of switches for physicians of different ages.

In contrast to other studies of movers that use regional variation in intensity, our measure of origin-group intensity may reflect the physician's own behavior. Specifically, the switching physician may influence the practice intensity of her peers in the origin group. Given that our main explanatory variable is the difference in treatment intensity across the destination and origin groups, we run a number of checks to ensure that this potential source of endogeneity is not driving our results. First, we estimate models where we use the destination-group intensity as the main explanatory variable of interest rather than the change in group intensity. This forsakes the useful variation in the shock to group intensity that comes from variation in origin-group levels, but it relies on a potentially more-exogenous measure of the shock to the practice environment. Second, we flexibly control for origin-group intensity by reporting results for different quantiles of origin-group intensity.

An interpretation issue arises from the fact that more-intensive groups may differ along other dimensions as well, such as physician training and beliefs. We will directly test the influence of a characteristic that we can easily measure and is policy relevant: group size. In the end, the physician's change in treatment intensity could stem from these other factors, and we view our estimates as a test of whether group affiliation matters for treatment intensity and health outcomes.

# 4 Data and Sample Description

## 4.1 Data

Our primary source of data is Medicare Claims data from 2008 to 2016. To measure physician treatment intensity, we rely on claims for a 20% random sample of beneficiaries in the Carrier file where payments to physicians are recorded. Payments for physician fees are made on a fee-for-service basis; physicians can increase their reimbursement from Medicare by increasing the services they provide to patients, by selecting more expensive services, or by treating more patients. We will test for the influence of group intensity on all of these measures. The claim includes lines-of-service coded using the Healthcare Common Procedure Coding System (HCPCS), analogous to commercial Current Procedural Terminology (CPT) codes. With these codes we investigate whether the types of claims change after a switch and begin to consider changes in coding behavior.

Importantly, these data also include a billing identifier (ID), the (deidentified) Tax Identitification Number, and we identify groups based on physicians billing under the same ID. Several other papers have used these billing IDs to identify groups (Austin and Baker, 2015; Baker et al., 2014b; Ketcham et al., 2007; Welch et al., 2013). There are at least two potential limitations when relying on such a billing ID to characterize the environmental intensity of a physician's group. First, the ID may represent a much larger organization, and the other physicians in the group may not exert as much influence as those working in the same team as a smaller unit, such as within a clinic (Welch et al., 2013); alternatively, a particularly large group may bill under more than one ID (Capps et al., 2018). However, for summary measures of group treatment intensity as measured by Medicare spending, the group billing ID is close to what we are seeking to characterize. Second, given that the Carrier file represents a 20% random sample of beneficiaries, we are likely not capturing all physicians associated with a given group, which introduces measurement error as well.

We measure each group's intensity as the average per-physician total reimbursement per quarter for all physicians in the group except the switching physician, across the four quarters prior to the switch quarter, weighted by the number of patients that a physician treats in each quarter.[5] In this way, physicians' contributions to the intensity of the group environment are representative of how active they are in the group. Our results are robust to non-weighted measures of average intensity. These data also include patient characteristics, including age, race, and sex, and because the data are longitudinal we are able to observe claims for the beneficiary over the year prior to an admission.

We merge the physicians' claims records to the 100% Inpatient files to identify the hospital associated with a given stay and the hospitals where physicians work. We also use the Inpatient files to record the admission and discharge dates associated with that hospital stay in order to calculate length of stay, 30-day readmission rates, and the number of major procedures associated with a given stay.[6] These data also provide additional information on diagnoses.

We use the National Plan and Provider Enumeration System (NPPES) dataset to obtain additional information about physicians, including gender and specialty, and to differentiate between physicians and other medical professionals. We also use the CMS Physician Compare database to obtain information on physician experience (in years), based on the year they graduated from medical school. We match 85% of our final treated sample of physicians to this database. In robustness checks where we search for

[5]We use the claim's summary payment amount measure, available in the Carrier files, which is the sum of payments made by CMS to the physician and the beneficiary. Beneficiary payments tend to be negligible on average ($< 0.1\%$ of the total payment), and thus we take these payments to characterize the amount a physician receives from CMS.

[6]We merge Carrier and Inpatient claim records based on deidentified patient ID and dates of service. According to conversations with the Research Data Assistance Center, an advantage of this approach over relying solely on the place-of-service codes in the Carrier files is that it more accurately captures hospital stays.

treatment effect heterogeneity by years of experience, we focus our analysis on this sub-sample. Last, we use American Hospital Association survey data to identify whether a hospital is a general acute care hospital.

## 4.2 Outcomes

Our outcomes are estimated at the physician-quarter level and are intended to capture measures of treatment intensity and quality of care. Treatment intensity includes three summary measures of quarterly intensity and three normalized measures. The quarterly summary measures are total Medicare reimbursement that the physician received (which is also how we characterize group intensity), total patients treated in the quarter, and total line items filed (each with a HCPCS code) to measure the number of services provided in the quarter. The normalized measures include reimbursement per patient, reimbursement per HCPCS, and HCPCS per patient. With these measures, we test whether a switching physician treats a given patient more intensively or performs more expensive procedures following a switch to a more-intensive group.

We also estimate effects on two hospitalization-level measures of treatment intensity. First, using the Inpatient claims, we observe the change in the average number of major procedures performed in each quarter among patients treated by each physician. Because these procedures are linked to the entire hospitalization, and thus not necessarily attributable to the switching physician, we include these procedures as a representation of broader treatment intensity (attributing procedures to *all* physicians who had corresponding Carrier claims associated with that hospitalization). We also include average length of stay (per patient-quarter) as a per-patient measure of treatment intensity because it is positively correlated with the quantity of treatment provided.

Next, we include several measures intended to capture changes in the quality of care provided. First, we calculate a physician's 30-day readmission rate as the share of all patients the physician treated in a given quarter who had a readmission within 30 days of the discharge date. Second, we calculate a physician's 30- and 365-day mortality rates as the share of hospitalizations in which the patient died within 30 or 365 days of admission. The mortality measures stem from vital statistics records, so we observe mortality regardless of whether it occurs in a hospital or not.

These measures are commonly used to evaluate quality of care provided. Thirty-day readmission is used by CMS as a quality measure.[7] The 30-day mortality rate in particular is included in Hospital

[7]Note that we do not differentiate between unplanned 30-day readmissions, which are penalized by CMS, and planned

Compare data as a measure of hospital quality (Doyle et al., 2019). Note that we are attributing these readmission and mortality rates to physicians who are *not necessarily* listed as the attending physician on the hospitalization record, but instead have a corresponding carrier line item during the hospitalization. This approach allows us to estimate a readmission and mortality measure for each physician in our sample, though potentially deviates from more conventional approaches of attributing readmissions/mortality to the attending physician on record.

Tables A.1 and A.2 report summary statistics of the level and log versions of our outcomes, respectively, as well as statistics of the main scaling variables.

## 4.3 Sample Construction

Our study sample is comprised of two physician cohorts: physicians who switch groups ("switchers"), and physicians who never switch groups ("non-switchers"), whose primary function is to increase the precision with which we can estimate and control for hospital- and calendar-level secular tends. We make a number of sample restrictions to implement our estimation strategy, as shown in Table A.3. Because we examine the effect of group environment on physician intensity, we attribute physicians to exactly one group per quarter, where group assignment is determined by the billing identifier associated with at least 90% of their claim line items (represented by HCPCS) for which they file for reimbursement in that particular quarter.[8] In the quarter prior to a switch, physicians are in hospitals with an average of 14 groups (SD: 11) that have at least one internist member, and an average of 8 (SD: 9) groups comprised only of internists in that particular quarter. Of the 553,721 physicians in our starting sample, we were able to attribute 552,420 to one group per quarter.

We define a switching episode for each switcher physician by identifying a period of at least nine consecutive quarters during which the physician belongs to a given origin group for at least four consecutive quarters, switches to a destination group in a "switch quarter," and belongs to that destination group for at least four consecutive quarters thereafter. By this definition, switching physicians can have multiple episodes. We observe 72,426 physicians who ever switch, and 83,870 switching episodes; each switching physician is associated with an average of 1.16 (SD: 0.40) episodes. Figure A.2 plots the share of HCPCS associated with a given origin or destination group for physicians in our final sample, in the

readmissions in order to measure total resources used.

[8]On average, physicians associate 92% (SD: 22%) of their HCPCS with a particular billing ID in any given quarter. In approximately 3% of treated physician-quarters outside of the switch quarter, physicians with an internal medicine specialty are attributed to groups associated with less than 90% of HCPCS in that quarter because they had more than 90% of HCPCS associated with a single group in the surrounding quarters.

quarters relative to the switch. As is evident from the figure, there is a transition quarter at the time of the switch ($q = 0$), during which physicians transition out of their origin group (average share of HCPCS: 0.75) to the destination group.

Non-switcher physicians include any physician who is observed to be attributed to only one group throughout the study period, which we similarly refer to as their "episode" for sake of consistency. By this definition, we flag 321,963 never-switching physicians, included for an average of 15.9 (SD: 13.3) quarters during our study period. The other physicians who are dropped at this step were in multiple groups but did not meet our nine consecutive quarter restriction (to be included in our switcher cohort).

To focus on within-hospital variation in treatment intensity, we further restrict the sample to physicians practicing within one hospital during each episode. Note that switcher physicians may switch hospitals at some point during the study period, as long as it does not occur contemporaneously with a group-switch episode. We attribute each physician to exactly one hospital per quarter by assigning them to the hospital associated with the plurality of their HCPCS in a given quarter.[9] On average, physicians associate 62% (SD: 40%) of their HCPCS with a particular hospital in any given quarter.

In order to be included in the final sample, both the origin and destination groups must exist in the four quarters prior to the physician's switch, as this is the relevant time period for identifying the change in environmental intensity. Additionally, in order to calculate the change in environmental intensity, which is estimated based on the average intensity of the *other* physicians in the group, at least one other physician (in addition to the switching physician) must belong to the origin groups. This restriction limits the analysis in two ways: First, we cannot observe origin groups in which the switching physician was the solo practitioner. Instead, we evaluate how effects vary by size of the origin and destination groups to see whether there is a relationship between origin-group size and our main results. Second, destination groups that do not exist in the pre-switch period are excluded from the sample, which excludes any group that forms in the post-switch period. This restriction focuses our analysis on changes in a physician's own intensity level due to a change in group intensity that arises from already-established group environments. Finally, to calculate group intensity (and to mitigate measurement error), we require that each origin and destination group treat at least 10 patients per quarter. We show that our results are robust to different cutoffs.

After imposing these restrictions, we have 162,433 non-switching physicians, 30,887 of whom have a

[9] In the instance of a tie (i.e. a quarter in which the physician has equal HCPCS across multiple hospitals), we default to the general acute care hospital, and remaining ties are broken at random; these ties occur for approximately 4% of physician-quarters.

specialty of internal medicine. Among switchers, we have 13,883 physicians (14,487 physician-episodes), including 3,108 physicians with a specialty of internal medicine (3,242 physician-episodes). Table A.3 summarizes counts of physicians at each restriction step.

Because we only specify that switching physicians belong to an origin group for at least four quarters before the switch, and a destination group for at least four quarters after the switch, we have an unbalanced panel when we examine outcomes beyond those quarters. Physician fixed effects (detailed in our model in Section 3) control for any systematic, time-invariant differences between physicians that are in a given group for more than four quarters.

## 4.4 Summary Statistics

Table 1 reports summary statistics for the internal-medicine physician episodes that form the basis of our primary results; see Table A.4 for the full sample of physician episodes. Panel A describes different measures of treatment intensity. The first row shows our primary measure of the change in treatment intensity between origin and destination groups. Among switching physicians, the average difference in log-payments per physician between origin and destination groups over the quarters prior to the switch is -14 log points. The negative sign indicates that, on average, destination groups tend to be somewhat less intense than origin groups. For those who switch to a group that is relatively less intense, the average difference is -65 log points, while for those who switch to a more-intensive group, this average is 47 log points. Figure A.3 plots the distribution of the relative change in group intensity as a histogram, showing a standard deviation of 0.78.

In levels, average reimbursement-per-physician in the origin group for all internists is approximately $2300 in this 20% sample of Medicare beneficiaries in fee-for-service Medicare; for switchers, average origin group intensity is approximately $3600.[10] Physicians who leave origin groups for destination groups that are less intensive come from groups that have a relatively high pre-switch intensity ($4300), while physicians switching to more-intensive groups leave groups that are more similar in intensity to the overall average with a mean of $2700. Not surprisingly, the opposite trends are found for destination groups. Given that the level of intensity varies across physicians who choose different destination groups, our estimation relies on the fact that the trends in intensity across physicians who choose destination groups that vary relative to their origin groups are comparable in the absence of a switch, a feature that

[10] While we measure group intensity for switchers as a leave-one-out mean, we calculate group intensity for non-switchers as simply the overall average, inclusive of the index physician, for computational reasons.

we can evaluate in the periods before a switch.

Table 1 also shows intensity measures at the physician level to illustrate the types of physicians that switch. Average reimbursement among switchers in the pre-period is somewhat higher compared to all physicians, as their origin groups are also relatively more intensive. Among switchers, physicians tend to move to groups that are more similar to their pre-switch intensity: those moving to less-intensive groups tend to be approximately 11% less intensive than their origin-group peers prior to the switch, and similarly they are more intensive (approximately 5%) than their origin-group peers when they move to higher-intensity groups. The opposite is true when we compare their intensity to the intensity of their subsequent peers in the destination group: physicians switching to less-intensive groups tend to be more intensive than their destination group peers (by approximately 56%), while physicians switching to more-intensive groups tend to be less intensive than their new colleagues (by approximately 32%).

Note that whenever we cut the sample by the direction of the change in group intensity, the origin group is expected to be relatively more intensive when the physician is moving to a lower-intensity group and vice versa, so the comparisons of the switching physician to her former or future colleagues partly reflects this sample selection. In any event, these differences motivate our exploration below of results by different levels of origin-group treatment intensity, as well as different levels of physician treatment intensity prior to the switch.

Panel B reports patient characteristics, showing that the average age, race, and sex of patients is similar across these comparison groups. We also calculated a predicted mortality measure using patient demographics and diagnoses in Medicare claims data in the year prior to the hospitalization.[11] This measure is also similar across different types of moves. Panel C reports summary statistics for physician characteristics. Switchers are slightly less likely to be male than all physicians (69% versus 62% male, respectively). Switcher physicians have slightly fewer years of experience (21 years) compared to an overall average of 23. Physician characteristics are similar across different types of switches.

Panel D reports summary statistics for the physicians' groups to give a sense of scale. Switchers are in origin groups that treat 254 patients per quarter in this 20% sample, and they tend to have 93 physicians in the group. This is smaller than the group size for all physicians, which is just over 160. Switchers

[11] We calculate predicted 1-year mortality in the following steps. First, we use a linear model to estimate the relationship between an indicator for whether a patient died in 2012 or 2013 and patient age (in vigintiles), sex, race, and comorbidity indicators recorded in 2012, with 2012 being the midpoint of our study period. We exclude all patients treated by physicians in our final study sample from this analysis. Using the coefficients obtained from this regression, we predict 1-year mortality for each patient treated by a physician in our final study sample using comorbidity indicators from the year prior to the physician treating the patient.

tend to join larger groups, as indicated by the larger average size of destination groups, which tend to have approximately a third more physicians (around 130) as origin groups, and treat approximately 20% more patients than in origin groups. Average $\Delta_{size}$ (the change in log group size between destination and origin groups) reflects this general pattern in increasing size; the relative change in group size is consistently positive and similar for physicians regardless of the direction of their group-intensity switch. The movement towards larger groups, and how it corresponds to changes in group intensity, is something we explore in more detail below.

To complement this table, Figure 1 plots the distribution of group intensity (panel (a)) and physician intensity (panel (b)), as well as the relationship between the two. The standard deviations of overall and within-hospital group intensity are very similar, with variation of group intensity overall being slightly larger than within-hospital variation. Notably, the standard deviations are quite large; a one-standard deviation increase in overall group intensity is 1.01, and 0.85 for within-hospital group intensity. This large degree of variation in group intensity is interesting in its own right, and is also useful empirically for our identification strategy. As with group intensity, the standard deviations of physician intensity are quite large; a one standard deviation increase in physician intensity is greater than 100 log points for all measures of variation (overall, within-hospital, and within-groups). Notably, variation in intensity among internists is nearly as large as variation across specialties, which we report in Figure A.4. This large degree of variability is both remarkable and in line with prior literature (Epstein and Nicholson, 2009; Tsugawa et al., 2017).

Moreover, physician intensity is positively and strongly correlated with the intensity of their peer colleagues. Without any additional adjustments, panel (c) of Figure 1 shows that physicians who belong to an origin group that has a 100 log point higher group intensity (slightly larger than one standard deviation) has a 57 log point higher intensity level themselves.[12] This relationship does not account for any endogeneity that might be associated with both the group intensity and the physician's intensity, such as a physician's preference for practicing in a group similar to their preferred level of intensity, or features of more-intensive groups (such as increased physical or human capital) that influence a physician's intensity. We aim to control for these factors in our main results.

[12] For computational reasons, we calculated the intensity of the group a non-switcher belongs to as the leave-in mean; thus, own intensity is highly correlated with group intensity, particularly in smaller groups. Because we don't believe this to be informative, but rather reflective of a mechanical relationship that we avoid in our leave-one-out means, we do not include the relationship between own intensity and group intensity for non-switchers.

# 5 Main Results

## 5.1 Balance Checks

A primary concern when comparing treatment and health outcomes after a switch is that patients treated by switching physicians may differ in ways that are correlated with our outcomes. To begin to address this concern, we restrict our analysis to physicians who switch groups within a hospital and estimate our models using within-hospital variation, which in part controls for patient characteristics (assuming a time-invariant patient composition at the hospital level). Still, physician groups may treat different types of patients, raising endogeneity concerns.

Figure A.5 reports the event study estimates described above, where the outcome is a particular patient characteristic. Similar to Table 1, the plots suggest that these characteristics are balanced around the switch for physicians moving to groups that vary in their average intensity. We further test for balance in the clinical composition of patients by estimating predicted 1-year mortality (as a function of demographics and chronic conditions described above) for each patient treated by the physician, and plot the change in average predicted mortality across the switch in Figure A.6. We find no meaningful difference in predicted mortality, which we interpret as evidence against physicians treating demonstrably different patients following a group switch. Due to evidence that patient composition doesn't change along these dimensions, our main results do not control for patient characteristics, although we report results with these controls in the appendix (Table A.5).

## 5.2 Group Affiliation and Physician Reimbursement

Figure 2 presents the main event-study results. The horizontal axis represents the quarters to and from the time of the switch. The points represent the $\theta_q$s estimated in Equation 6: the difference in average quarterly spending in the quarters leading up to and lagging away from a switch, for switchers versus non-switchers, scaled by the difference in treatment intensity between the destination and origin groups. The left set of figures are for ln(Pmt), which is the log of Medicare reimbursements received by the physician. The right column of figures reports results for the related outcome of log reimbursement per patient. The first row shows that the relationship between physicians' treatment intensity and the eventual pre-post disparity in the group intensity measures is relatively flat and approximately 0 prior to the switch. We see a small jump in both measures of intensity in the quarter of the switch, followed by a substantial increase that remains steady for the following 10 quarters. Specifically, we observe that log reimbursement is relatively steady at approximately 0 log points prior to the switch quarter, and then rises to between 20 to 30 log points higher after the switch, staying relatively constant in the post-switch quarters. When we normalize payments by the number of patients seen, we observe a similar pattern with a sustained increase of between approximately 10 and 20 log points higher compared to the quarter just prior to the switch. Confidence intervals are similar when we bootstrap the standard errors (Figure A.7) to take into account that the measure of the change in group intensity is a generated regressor.

Panels (b) and (c) report the same event studies for physicians who join more (less) intensive groups. The results are somewhat noisier as expected given the smaller sample sizes, but the direction of the change in intensity is symmetric across the two types of moves. Physicians joining more-intensive groups increase their intensity by between 20 to 30 log points. For those who join less-intensive groups, the change in own intensity appears sustained at approximately 30 to 50 log points lower than pre-switch intensity.

Recent discussions surrounding contamination in the estimates of event-study models that include staggered events have prompted a re-examination of traditional estimation methods (Sun and Abraham, 2020). In particular, there is a concern that the event-study coefficients plotted need not represent an average of effects across the staggered events. To check this concern we estimated our model separately for each event period defined by the calendar quarter of the switch, and we compared our results to the average of these many event studies. We identify 28 switching cohorts with an average of 116 (SD: 51) physicians per cohort. In Figure 3, we plot the estimated $\theta_q$s from each cohort-specific estimation in gray, the average of these $\theta_q$s in blue, and the estimates from the main pooled estimation in red. The results indicate that our estimated effects closely approximate the average of effects across the cohorts.

Table 2 reports estimates from the simpler pre-post model given by Equation 7. Supporting the trends reported in the event study plots, we estimate that an increase in group intensity by 1 (or, 100 log points, which is similar to the 78 log-point standard deviation of our change in intensity measure) results in an approximately 29 log-point increase in a physician's own intensity. This is substantial, although significantly smaller than the raw correlation described above that implied an elasticity of closer to 0.6. Column (4) shows that the results for log payment per patient are slightly more than half of this magnitude: a similar change in group intensity results in an increase in payment per patient of approximately 19 log points.

## 5.3 Decomposition of Group and Physician Effects

As noted in Section 3.1 and in Finkelstein et al. (2016), in the AKM model we can decompose group variation into components that are attributed to the physician (such as preferences and beliefs about treatment) and groups (such as group management). Equation 5 shows that share of group variation attributed to the group effects is the slope of the relationship between $\Delta_{pmt}$ and the jump in physician's treatment intensity at the time of a switch. This can be flexibly estimated using a bin scatter plot of the average change in treatment intensity among switchers across bins of $\Delta_{pmt}$.

Figure 4 carries out this exercise using vigintiles of the treatment variable (the difference in treatment intensity across the destination and origin groups). Note that there are no controls in this specification; we simply bin the data according to vigintiles of $\Delta_{pmt}$ (i.e. change in group intensity) and regress average $\Delta_y$ (i.e. change in physician's intensity) against this change in pre-determined group intensity. One feature of the figure is that most of the points are below zero (up to approximately one standard deviation above the mean): when physicians switch they tend to be less intensive. The second feature is that the relationship with the change in group intensity is fairly linear, mirroring the results in Figure 2 that show the relative symmetry in the event studies across switches to more- or less-intensive groups. This is reassuring, as (1) it is consistent with the additively-separable AKM model, and (2) it suggests that the results are not driven by only a small portion of the distribution which might be related to shocks in the unobservables related to the switch. The coefficient of 0.18 suggests that approximately 18% of the observed variation in treatment intensity across groups can be attributed to a group-specific component, leaving 82% of the variation to be attributed to a physician-specific component among internal-medicine physicians.

The jump in intensity at the time of the switch from origin to destination group in the event study also provides an estimate of $S_{group}(g, g')$. We interpret the somewhat smaller magnitude of the slope in Figure 4 (0.18), compared to the size of the jump in Figure 2 (approximately 0.29), as evidence that the hospital-quarter controls in our main results account for general trends in intensity over this time period.

## 5.4 Group Affiliation and Other Measures of Intensity

While overall Medicare reimbursement and average Medicare reimbursement-per-patient are useful and policy-relevant measures of treatment intensity, we can begin to unpack the sources of these changes by considering related measures of intensity. Figure 5 reports results for six related measures. The first panel plots the change in (log) patients per quarter, and indicates that physicians who switch to more-intensive groups subsequently increase the number of patients they treat by between 5 to 10 log points. Similarly, the second panel plots the change in (log) HCPCS per quarter, and shows that switching to a more-intensive group (for a unit increase in our group-intensity measure) scales up the number of claim line-items filed in a given quarter by approximately 10 to 15 log points.

The plots in the second row indicate that both log HCPCS per patient and log reimbursement per HCPCS are scaled upwards by a physician switching to a more-intensive group, with relatively flat pre-trends. This suggests that physicians are billing for more line-items per patient, and that they are also adopting more intensive line-items per patient following a switch to a more-intensive group. The types of line items are explored in more detail below.

Finally, in the bottom row, we plot the post-switch change in log major procedures associated with a hospitalization, as well as log length of stay. As noted above, procedures are for the entire stay and may not have been influenced by the switching physician. This measure of intensity shows that the number of major procedures per patient increases by between 5 to 10 log points. The effect on average patient's length of stay is less precisely estimated and suggests a more muted effect.

To summarize these changes in a pre-post model, Panels A and B of Table 2 report the results. Recall that a typical increase in group intensity is on the order of 1 (100 log points), so we can simply read the coefficients listed in the table to interpret the effect of an increase in group intensity: when we compare the post period to the period prior to the move, we observe an 8 log-point increase in average patients treated per quarter and a 14 log-point increase in average HCPCS per quarter. For our additional intensive margin measures, we observe a 19 log-point increase in average quarterly payments per patient, a 13 log-point increase in payment per HCPCS, and a smaller 4 log-point increase in the number of HCPCS recorded per patient. The estimated effects on hospitalization-level measures similarly reflect the patterns observed in the event study plots: increasing group intensity by 1 increases major procedures and length of stay associated with a given hospitalization by 8 and 0.4 log points, respectively.

Taken together, the estimates suggest that physicians' group affiliation matters: when their group intensity increases, physicians increase their intensity in terms of the number of patients seen and the number of services provided. This continues to be found at the per-patient basis, with physicians performing tasks that have a higher reimbursement and a modest increase in the number of services provided per patient.

## 5.5 Treatment Intensity Robustness

Our measure of the change in group intensity is the difference between average intensity among physicians in the destination group minus the average intensity of other physicians in the origin group, both measured in the four quarters before the switch. However, if peers influence one another as we hypothesize, then the origin-group intensity will reflect the influence of the switching physician. Our summary statistics in Table 1 suggest that switchers to less-intensive groups are less intensive than their colleagues, and that switchers to more-intensive groups are more intensive than their colleagues. For example, if more-intensive switchers have a positive effect on their colleagues' intensity (i.e. "pull up" that intensity), then for those who switch to more-intensive groups, the change would have been larger in the absence of the switching-physician's influence. This would suggest that the jump that we find understates the effect of the group if we were able to control for the reflection problem.

We implement a number of checks to help address this concern. First, we estimate effects for different levels of origin-group intensity and find similar results regardless of the quartile of origin group intensity (Figure A.8). Second, we estimate our model across quartiles of physicians' pre-switch intensity, finding relatively similar effects for those that are less- or more-intensive in the pre-switch period (Figure A.9). Third, the relatively symmetric results in Figure 2 help rule out an explanation where more-intensive physicians have a significantly stronger influence on others compared to less-intensive physicians. Fourth, we expect that a physician would have a larger influence on her peers in small groups, yet we find that if anything, estimated effects are smallest in smaller origin groups. Overall, we do not see a clear pattern in results across group-size quartiles (Figure A.10).

In addition, results are very similar when we only use the destination-group intensity as the treatment of interest rather than the change in intensity (Figure A.11). While we do not rule out cross-group influences prior to the switch given our within-hospital design, these robustness results suggest that a reflection problem in the origin group, which would tend to lead our main result to underestimate the group influence, is not affecting the main results.

## 5.6 Effects on Health Outcomes

The results above suggest that most of the difference in treatment intensity across physicians within a hospital comes from characteristics of the physicians themselves, but we also find a substantial influence of group affiliation on treatment intensity. This influence results in meaningful changes in treatment intensity after a switch, in part because groups differ so widely in their intensity levels. To evaluate the welfare implications, we consider health outcomes.

Figure 6 reports the results for readmissions and mortality. Despite a sudden and sustained change in Medicare billing and procedures performed, the figures show a relatively flat relationship between the timing of the switch and patient outcomes. The mortality coefficients are both positive and negative in the post period, and within a fairly narrow range.

To gain precision, we again use a pre-post model and report the results in Table 2. Here we see that after the switch, a one-unit increase in group intensity is associated with a 0.2% increase in 30-day readmission, or about 1% of the mean of 24%. For 30-day mortality, a typical increase in group intensity (a change of 1) is associated with a 0.3% increase in mortality compared to a mean of approximately 10% (i.e. approximately 3% of the mean). For one-year mortality, the coefficient represents a 0.8% increase compared to a mean of 30% (approximately 2% of the mean). The estimates are reasonably precise: the lower bound of the confidence interval for one-year mortality is 0.001, which is 0.4% of the mean. With small, positive point estimates on both readmissions and mortality, it appears that the change in treatment intensity is not associated with improved outcomes for patients.

Perhaps the lack of an effect on major outcomes is not surprising given the magnitudes of the spending differences, which are fairly small on their own.[13] That said, the estimated increases in the price per claim and the number of line-items billed are not trivial. Another reason we may not see the change in treatment intensity translate to these major health outcomes is that the physician is only one of many who might treat any given patient in the hospital, so her own effect may be diluted. Rather, we view the results as consistent with group affiliation affecting treatment intensity resulting in a modest effect on Medicare spending with no detectable effect of a change in health outcomes.

## 5.7 Effects of Group Size

When physicians switch groups, it's likely that group intensity is not the only component of their practice environment that changes. One particularly policy-relevant characteristic of the physician group is its size, as measured by the number of physicians in the group. As noted above, there is an increasing trend towards larger physician groups via consolidation. Table 1 suggests that, consistent with this overall

[13] Average log-billing per patient based on Table 2 is 5.267, or approximately $194, and our estimates suggest that a switch to a higher-spending group would increase spending by approximately 14%, or approximately $37 per patient.

trend, switchers are on average moving to larger groups. A natural follow-up question is how a change in group size might influence a physician's intensity. In Figure A.13, we plot the relationship between physician intensity and group intensity. We find a small but significant negative relationship, suggesting that physicians in larger groups are less intense. This relationship is also reflected in group intensity; panel (b) shows that as group size increases, group intensity decreases.

To investigate the causal relationship between the number of physicians in a group and physician intensity, we estimate Equation 6 scaling the leads and lags by log change in group size instead of log change in intensity. The distribution of $\Delta_{size}$ is plotted in Figure A.12, revealing a mean of 47 log points and a standard deviation of 157 log points.

Figures 7 and 8 are analogous to Figures 2 and 6, but now use the change in group size rather than the change in group treatment intensity as a measure of changing group environment. Figure 7 shows that our main measures of treatment intensity are relatively stable prior to the switch, and then drop at the time of the switch. Notably, the magnitudes are smaller and less precisely estimated than in our main change-in-intensity specification. For such a change, the event-study figures suggest that intensity falls by between 2 to 7 log points. Similar to our estimates using log change in intensity, Figure 8 reports no meaningful change in our quality-of-care measures across the switch when we scale leads and lags by log change in group size.

Panel A of Table 3 shows that for a 100 log point increase in group size, payments fall by approximately 5%, patients treated fall by 3%, and the number of HCPCS fall by about 4%. Panel B of Table 3 reports a statistically-insignificant 1% reduction in payments per patient and a decrease in payment per HCPCS of approximately 0.5%. We observe a marginally significant 0.7% drop in HCPCS per patient, and a significant 2.6% decrease in major procedures per quarter. Figure A.14 reports the estimates of the $\theta_q$s for our other intensive and hospital-level measures of intensity. Estimates are relatively noisy, with some indication of a sudden decrease in log HCPCS following the switch. These estimates offer suggestive evidence that treatment intensity decreases as physicians move to larger groups.

With so much attention on group size increasing over time, these results suggest that a physician's own treatment intensity does not rise with larger groups, with some evidence that intensity falls. Meanwhile, Panel C of Table 3 reports small, positive point estimates of the effects of a change in group size on readmissions and mortality, with these effects not statistically significant. This is consistent with Figure 8, which does not show a sustained change in these outcomes before and after the switch.

To explore how a change in group intensity might be mediated by a change in group size, Figure A.15

plots the $\theta_q$s obtained from estimating our primary specification separately by quartile of the change in (log) group size. We find relatively consistent estimates of a jump in physician's own intensity post-switch, with a smaller magnitude estimated for physicians switching to larger groups in the largest quartile of $\Delta_{size}$. We also estimate our event study allowing for heterogeneity in treatment effects by the size quartile of the destination group (Figure A.16); we find a relatively uniform effect across destination group size quartiles.

## 5.8 Heterogeneity Across Specialties

The main results focus on internal medicine physicians to control for the type of care provided, and internal medicine is the most common specialty among the switching physicians in our sample. We also explore whether groups affect treatment intensity and health outcomes among other types of physicians.

When we estimate our model on the pooled sample of all physicians, we find similar patterns, but the magnitudes of the effects are smaller (Figure A.17). We also observe similar symmetry in the effect among physicians who switch to relatively more- versus less-intensive groups (panels (b) and (c)). Similar to our observations among internists, Figure A.18 indicates no meaningful effect of a switch to a more-intensive group on measures of quality of care. We confirm that the effects we observe in the pooled sample are not solely being driven by internists by exploring estimates using a sample of all non-internist physicians, finding that the effect of group intensity on physician's own intensity has a similar pattern in log payments and payments per patient, albeit of smaller magnitude, as well as very little to no effect on quality of care outcomes among non-internal medicine specialties (Figure A.19). When we estimate our model on another common specialty (cardiologists), we find similar effects of group intensity on our outcomes of interest. However, the effects are smaller and noisier, potentially due to the smaller sample size. Tables A.6 and A.7 report the pre-post estimates for all specialties and non-internists, respectively. The group-size results are also similar to the main results.

Finally, we implement our decomposition exercise and plot average $\Delta_y$ for all physicians against the associated $\Delta_{pmt}$ vigintiles. Figure A.20 reports a linear relationship, with a coefficient of 0.11, suggesting that approximately 11% of the observed variation in treatment intensity across groups can be attributed to a group-specific component. It appears that the influence of a change in group environmental intensity applies to a range of physicians, with larger effects found for internists.

### 5.9 Additional Robustness Checks

The main identifying assumption is that there is not a contemporaneous shock at the time of a switch, such as a change in patient characteristics or preferred practice style that might also explain a change in practice intensity. For example, physicians switching at different points in their career could point to changing priorities with regards to practice intensity. One way to test this identifying assumption is to compare the effects measured for physicians who have different levels of experience. If, for example, physicians close to retirement are switching to less-intensive groups at the same time as they are changing their preferred practice style (i.e. "ramping down"), we might be concerned about bias in our estimates.

Figure A.21 plots the main coefficients of interest ($\theta_q$s) by quartiles of physicians' years of experience. All four quartiles show a relatively flat pre-trend followed by a sustained increase in intensity, including for those with the most experience and for whom we might expect that a group switch may be part of an effort to scale back on work. We view this evidence as supportive of our main identifying assumption.

We also conduct a more targeted examination of quality measures by exploring the effect of a switch on mortality among patients aged 85 and older who have higher mortality rates. Figure A.22 plots the $\theta_q$s obtained from estimating the model on the share of patients older than 85 who die within 30 days and one year. As in our main mortality measures, we observe no meaningful change in mortality surrounding the group switch, reinforcing the primary findings of a largely null effect of increased intensity on welfare gains.

Finally, we test the robustness of our 10-patient-per-quarter minimum inclusion criteria for switcher origin and destination groups. Specifically, we estimate Equation 6 using 5-patient and 20-patient cutoffs. As Figure A.23 shows, the estimated effects of a group switch are largely unchanged, which we interpret as evidence that our results are robust to measurement error in how we capture group intensity.

## 6 Coding Intensity

One mechanism that could drive an increase in billing is through changes in coding behavior. For example, Dafny (2005) documented hospitals "upcoding" patients to higher-paying diagnosis-related groups (DRGs) following a policy that changed reimbursement for certain DRGs. The most common types of HCPCS among internists are evaluation and management (E&M) visits; they account for the vast majority of the claim line-items we observe. A feature of these types of HCPCS is that they have different billing levels depending on patient complexity, and this complexity requires documentation.

Some groups may be more efficient in coding visits to increase revenue.

E&M visits in an inpatient setting have three levels of increasing intensity. Table A.8 reports CPT codes and associated average reimbursement associated with each code. In our data, a physician who conducts a level one inpatient initial E&M visit (of approximately 30 minutes duration; CPT code 99221) is reimbursed $97.40 on average. In comparison, a physician who conducts a level three inpatient initial E&M visit (of approximately 70 minutes duration; CPT code 99223) is reimbursed $194.89 on average.[14]

Figure 9 plots the change in log volume of these E&M visits, by level, following a switch, scaled by the change in group intensity between the destination and origin group as in our main specification (Equation 6). We observe no meaningful change in level 1 visits. We observe a significant increase in level 2 and 3 E&M visits, with a slightly larger increase in level 3 visits. This suggests that in addition to performing more HCPCS, physicians are billing at a higher intensity as they switch to more-intensive groups. In the context of the balanced patient characteristics across the switch (Figures A.5 and A.6), this further suggests that the increased billing intensity is not reflective of treating patients with increased clinical complexity.

Figure A.24 plots the change in E&M visits, by level, separately for physicians switching to less (column (a)) and more (column (b)) intensive groups. The top row shows little to no change in level 1 E&M visits as physicians switch groups. The middle and bottom rows, in contrast, show clear and symmetric changes in level 2 and level 3 visits, respectively. This effect is most symmetric (and largest) for level 3 visits, such that physicians switching to a group that is 100 log-point more (less) intense than their origin group increase (reduce) level 3 visits by between 20 to 40 log points.

In terms of the most common billing codes, E&M visits, the results suggest that either physicians are spending more time with patients or they are coding in a way that reflects more time spent. Either way, the higher-priced line items in the physician's claim do not appear to be the result of a change in patient complexity, and we do not find an improvement in patient health outcomes. This suggests that this group-induced change in intensity may not be productive.

Additionally, we considered two other measures of coding changes: the number of distinct diagnoses recorded per patient, and the use of diagnoses that signal higher patient complexity.[15] In both exercises, we do not find compelling evidence of changes in these coding practices associated with switching to

[14]See https://emuniversity.com/Page2.html and https://emuniversity.com/Page4.html for more details.

[15]As a proxy for diagnoses that signal higher complexity, we used a mapping of diagnosis codes to diagnostic-related groups that are characterized as conditions with complications or comorbidities; or with major complications or comorbidities.

more intensive groups. Such results are reassuring that the patient complexity does not appear to be changing with such a switch.

# 7 Conclusion

Most physicians work in group practices in an effort to reduce their own financial burden, legal exposure, and resource requirements. In this environment, a natural question that arises is, how does group affiliation affect a physician's own practice and, ultimately, how does this affect patient health outcomes?

We find that when physicians switch groups within the same hospital, their treatment intensity moves in the direction of the group they join. Among internists, the elasticity of own intensity with respect to group intensity is approximately 0.3. The results suggest that while most of the cross-group variation in intensity is due to physician factors, group factors explain 20-30%. The share attributed to group effects is lower for other specialties. While the change in treatment intensity scales with the change in group intensity, we find no corresponding, sustained change in patient health outcomes measured by readmissions and mortality.

The results have a number of limitations. First, we estimate the influence of group intensity on switchers, who may be more (or less) influenced by group affiliation compared to those who remain in the same group. Second, physician preferences could change at the same time as a switch, such as instances when a change in physician circumstances leads them to make a move. If this is the case, then the sudden and permanent change evident in the event studies suggests that physicians have to wait until they make the move before they can realize their new level of preferred treatment intensity, which would still be consistent with the group exerting a considerable influence in how treatment intensity is determined.

Third, given our use of a leave-out estimator of group intensity, we cannot estimate our model on physicians who switch from a solo practice to a larger group. However, the vast majority of physicians practice in groups, and this share is increasing over time. In addition, our results are similar when we consider different sizes of groups that physicians join.

Fourth, our results speak specifically to group influence in an inpatient setting, where there may be less physician discretion for treatment. This approach controls for time-invariant characteristics of the practice setting, but the results are less likely to apply to the outpatient setting, a subject for future research in this area. Fifth, our identification and characterization of groups and their intensity solely reflects physicians' treatment of Medicare beneficiaries. Medicare patients were found to comprise the plurality of direct hospital employee's patient panels in 2016 (Gillis, 2017), and are relevant for Medicare payment policy.

Despite these limitations, the results suggest that group affiliation has a sizeable effect on physician treatment intensity. This helps inform the sources of the remarkable amount of variation across physicians, even those practicing similar roles in the same practice setting. As a result, efforts to restrain healthcare spending may benefit from changing incentives and constraints at the level of the physician group.

# References


John M. Abowd, Francis Kramarz, and David N. Margolis. High wage workers and high wage firms. *Econometrica*, 67(2):251–333, 1999. doi: https://doi.org/10.1111/1468-0262.00020. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/1468-0262.00020.

Leila Agha, Keith Marzilli Ericson, and Xiaoxi Zhao. The impact of organizational boundaries on healthcare coordination and utilization. Working Paper 28179, National Bureau of Economic Research, December 2020. URL http://www.nber.org/papers/w28179.

Diane Alexander. Does physician pay affect procedure choice and patient health? evidence from medicaid c-section use. Working Paper WP-2017-7, FRB of Chicago, July 2015. URL https://ssrn.com/abstract=2984189.

Diane Alexander and Molly Schnell. The impacts of physician payments on patient access, use, and health. Working Paper 26095, National Bureau of Economic Research, July 2019. URL http://www.nber.org/papers/w26095.

Daniel R. Austin and Laurence C. Baker. Less physician practice competition is associated with higher prices paid for common procedures. *Health Affairs*, 34(10):1753–1760, 2015. doi: 10.1377/hlthaff.2015.0412. URL https://doi.org/10.1377/hlthaff.2015.0412.

Laurence C. Baker, M. Kate Bundorf, and Daniel P. Kessler. Patients' preferences explain a small but significant share of regional variation in medicare spending. *Health Affairs*, 33(6):957–963, 2014a. doi: 10.1377/hlthaff.2013.1184. URL https://doi.org/10.1377/hlthaff.2013.1184. PMID: 24889944.

Laurence C. Baker, M. Kate Bundorf, Anne B. Royalty, and Zachary Levin. Physician practice competition and prices paid by private insurers for office visits. *JAMA*, 312(16):1653–1662, 10 2014b. ISSN 0098-7484. doi: 10.1001/jama.2014.10921. URL https://doi.org/10.1001/jama.2014.10921.

Amber E. Barnato, M. Brooke Herndon, Denise L. Anthony, Patricia M. Gallagher, Jonathan S. Skinner, Julie P. W. Bynum, and Elliot S. Fisher. Are regional variations in end-of-life care intensity explained by patient preferences? a study of the us medicare population. *Med Care*, 45(5):386–393, May 2007. doi: 10.1097/01.mlr.0000255248.79308.41.

Ernst R. Berndt, Robert S. Gibbons, Anton Kolotilin, and Anna Levine Taub. The heterogeneity of concentrated prescribing behavior: Theory and evidence from antipsychotics. *Journal of Health Economics*, 40:26–39, 2015. doi: 10.1016/j.jhealeco.2014.11.003.

Marika Cabral, Colleen Carey, and Sarah Miller. The impact of provider payments on health care utilization: Evidence from medicare and medicaid. *NBER Working Paper*, (29471), 2021.

Cory Capps, David Dranove, and Christopher Ody. Physician practice consolidation driven by small acquisitions, so antitrust agencies have few tools to intervene. *Health Affairs*, 36(9):1556–1563, 2017. doi: 10.1377/hlthaff.2017.0054. URL https://doi.org/10.1377/hlthaff.2017.0054. PMID: 28874481.

Cory Capps, David Dranove, and Christopher Ody. The effect of hospital acquisitions of physician practices on prices and spending. *Journal of Health Economics*, 59:139 – 152, 2018. ISSN 0167-6296. doi: https://doi.org/10.1016/j.jhealeco.2018.04.001. URL http://www.sciencedirect.com/science/article/pii/S016762961730485X.

Lawrence P. Casalino, Michael F. Pesko, Andrew M. Ryan, Jayme L. Mendelsohn, Kennon R. Copeland, Patricia Pamela Ramsay, Xuming Sun, Diane R. Rittenhouse, and Stephen M. Shortell. Small primary care physician practices have low rates of preventable hospital admissions. *Health Affairs*, 33(9):1680–1688, 2014. doi: 10.1377/hlthaff.2014.0434. URL https://doi.org/10.1377/hlthaff.2014.0434. PMID: 25122562.

David C. Chan. Teamwork and moral hazard: Evidence from the emergency department. *Journal of Political Economy*, 124(3):734–770, 2016. doi: 10.1086/685910. URL https://doi.org/10.1086/685910.

David C. Chan. Influence and information in team decisions: Evidence from medical residency. *merican Economic Journal: Economic Policy*, 13(1):106–137, 2021. doi: 10.1257/pol.20180501.

Cindy Y. Chang and Ziad Obermeyer. Association of Clinical Characteristics With Variation in Emergency Physician Preferences for Patients. *JAMA Network Open*, 3(1):e1919607–e1919607, 01 2020. ISSN 2574-3805. doi: 10.1001/jamanetworkopen.2019.19607. URL https://doi.org/10.1001/jamanetworkopen.2019.19607.

Jeffrey Clemens and Joshua D. Gottlieb. Do physicians' financial incentives affect medical treatment and patient health? *The American Economic Review*, 104(4):1320–1349, 2014. ISSN 00028282. URL http://www.jstor.org/stable/42920741.

David Cutler, Jonathan S. Skinner, Ariel Dora Stern, and David Wennberg. Physician beliefs and patient preferences: A new look at regional variation in health care spending. *American Economic Journal: Economic Policy*, 11(1):192–221, February 2019. doi: 10.1257/pol.20150421. URL https://www.aeaweb.org/articles?id=10.1257/pol.20150421.

Leemore S. Dafny. How do hospitals respond to price changes? *American Economic Review*, 95(5):1525–1547, December 2005. doi: 10.1257/000282805775014236. URL https://www.aeaweb.org/articles?id=10.1257/000282805775014236.

Joseph Doyle, John Graves, and Jonathan Gruber. Evaluating Measures of Hospital Quality: Evidence from Ambulance Referral Patterns. *The Review of Economics and Statistics*, 101(5):841–852, 12 2019. ISSN 0034-6535. doi: 10.1162/rest_a_00804. URL https://doi.org/10.1162/rest_a_00804.

Randall P. Ellis and Thomas G. McGuire. Provider behavior under prospective reimbursement: Cost sharing and supply. *Journal of Health Economics*, 5:129–151, 1986. doi: https://doi.org/10.1016/0167-6296(86)90002-0.

Andrew J. Epstein and Sean Nicholson. The formation and evolution of physician treatment styles: An application to cesarean sections. *Journal of Health Economics*, 28(6):1126 – 1140, 2009. ISSN 0167-6296. doi: https://doi.org/10.1016/j.jhealeco.2009.08.003. URL http://www.sciencedirect.com/science/article/pii/S0167629609000770.

Andrew J. Epstein, Jonathan D. Ketcham, and Sean Nicholson. Specialization and matching in professional services firms. *The RAND Journal of Economics*, 41(4):811–834, 2010. ISSN 07416261. URL http://www.jstor.org/stable/25746055.

Amy Finkelstein, Matthew Gentzkow, and Heidi Williams. Sources of Geographic Variation in Health Care: Evidence From Patient Migration. *The Quarterly Journal of Economics*, 131(4):1681–1726, 07 2016. ISSN 0033-5533. doi: 10.1093/qje/qjw023. URL https://doi.org/10.1093/qje/qjw023.

Elliott S. Fisher, David E. Wennberg, Therese A. Stukel, Daniel J. Gottlieb, F.L. Lucas, and Etoile L. Pinder. The implications of regional variations in medicare spending. part 1: The content, quality, and accessibility of care. *Annals of Internal Medicine*, 138(4):273–287, 2003a. doi: 10.7326/0003-4819-138-4-200302180-00006. URL https://www.acpjournals.org/doi/abs/10.7326/0003-4819-138-4-200302180-00006. PMID: 12585825.

Elliott S. Fisher, David E. Wennberg, Therese A. Stukel, Daniel J. Gottlieb, F.L. Lucas, and Etoile L. Pinder. The implications of regional variations in medicare spending. part 2: Health outcomes and satisfaction with care. *Annals of Internal Medicine*, 138(4):288–298, 2003b. doi: 10.7326/0003-4819-138-4-200302180-00007. URL https://www.acpjournals.org/doi/abs/10.7326/0003-4819-138-4-200302180-00007. PMID: 12585826.

Kurt D. Gillis. Physicians' patient mix —a snapshot from the 2016 benchmark survey and changes associated with the aca. Technical report, American Medical Association, 2017. URL https://www.ama-assn.org/sites/ama-assn.org/files/corp/media-browser/public/health-policy/PRP-2017-physician-benchmark-survey-patient-mix.pdf.

Gardiner Harris. More doctors giving up private practices. *New York Times*, Mar 2010. URL https://www.nytimes.com/2010/03/26/health/policy/26docs.html.

J Heeringa, A Mutti, MF Furukawa, A Lechner, KA Maurer, and E. Rich. Horizontal and vertical integration of health care providers: A framework for understanding various provider organizational structures. *Journal of Integrated Care*, 1(20):1–10, 2020. doi: 10.5334/ijic.4635.

Carol K. Kane. Updated data on physician practice arrangements: Physician ownership drops below 50 percent. Policy research perspectives, American Medical Association, 2017.

Jonathan D. Ketcham, Laurence C. Baker, and Donna MacIsaac. Physician practice size and variations in treatments and outcomes: Evidence from medicare patients with ami. *Health Affairs*, 26(1):195–205, 2007. doi: 10.1377/hlthaff.26.1.195. URL https://doi.org/10.1377/hlthaff.26.1.195.

Charles F. Manski. Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, 60(3):531–542, 1993. ISSN 00346527, 1467937X. URL http://www.jstor.org/stable/2298123.

Thomas G McGuire and Mark V Pauly. Physician response to fee changes with multiple payers. *Journal of Health Economics*, 10(4):385 – 410, 1991. ISSN 0167-6296. doi: https://doi.org/10.1016/0167-6296(91)90022-F. URL http://www.sciencedirect.com/science/article/pii/016762969190022F.

J. Michael McWilliams, Michael E. Chernew, Alan M. Zaslavsky, Pasha Hamed, and Bruce E. Landon. Delivery System Integration and Health Care Spending and Quality for Medicare Beneficiaries. *JAMA Internal Medicine*, 173(15):1447–1456, 08 2013. ISSN 2168-6106. doi: 10.1001/jamainternmed.2013.6886. URL https://doi.org/10.1001/jamainternmed.2013.6886.

David Molitor. The evolution of physician practice styles: Evidence from cardiologist migration. *American Economic Journal: Economic Policy*, 10(1):326–56, February 2018. doi: 10.1257/pol.20160319. URL http://www.aeaweb.org/articles?id=10.1257/pol.20160319.

David B. Muhlestein and Nathan J. Smith. Physician consolidation: Rapid movement from small to large group practices, 2013–15. *Health Affairs*, 35(9):1638–1642, 2016. doi: 10.1377/hlthaff.2016.0130. URL https://doi.org/10.1377/hlthaff.2016.0130. PMID: 27605644.

Soroush Saghafian, Raha Imanirad, and Stephen J. Traub. Do physicians influence each other's performance? evidence from the emergency department. Working Paper RWP19-018, May 2019.

Lina D. Song, Soroush Saghafian, Joseph P. Newhouse, Mary Beth Landrum, and John Hsu. The impact of vertical integration on physician behavior and healthcare delivery: Evidence from gastroenterology practices. Working Paper RWP20-031, October 2020.

Liyang Sun and Sarah Abraham. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects, September 2020. URL http://economics.mit.edu/files/14964.

Yusuke Tsugawa, Ashish K. Jha, Joseph P. Newhouse, Alan M. Zaslavsky, and Anupam B. Jena. Variation in Physician Spending and Association With Patient Outcomes. *JAMA Internal Medicine*, 177(5):675–682, 05 2017. ISSN 2168-6106. doi: 10.1001/jamainternmed.2017.0059. URL https://doi.org/10.1001/jamainternmed.2017.0059.

W. Pete Welch, Alison Evans Cuellar, Sally C. Stearns, and Andrew B. Bindman. Proportion of physicians in large group practices continued to grow in 2009–11. *Health Affairs*, 32(9):1659–1666, 2013. doi: 10.1377/hlthaff.2012.1256. URL https://doi.org/10.1377/hlthaff.2012.1256. PMID: 24019373.

John Wennberg and Alan Gittelsohn. Small area variations in health care delivery. *Science*, 182(4117): 1102–1108, 1973. ISSN 0036-8075. doi: 10.1126/science.182.4117.1102. URL https://science.sciencemag.org/content/182/4117/1102.

Terry Zwiep, San (Hilalion) Ahn, Jamie Brehaut, Fady Balaa, Daniel I. McIsaac, Susan Rich, Tom Wallace, and Husein Moloo. Group practice impacts on patients, physicians and healthcare systems: a scoping review. *BMJ Open*, 11, 2021. doi: https://dx.doi.org/10.1136/bmjopen-2020-041579.

# 8 Tables and Figures

Table 1: Summary Statistics, Internal Medicine

| Measure | (1) All | (2) Switchers | (3) $\Delta_{pmt} < 0$ | (4) $\Delta_{pmt} > 0$ |
|---|---|---|---|---|
| A. Baseline treatment intensity | | | | |
| $\Delta_{pmt}$ | | -.14 | -.65 | .47 |
| Origin group reimb. per physician | 2300 | 3590 | 4311 | 2729 |
| Destination group reimb. per physician | | 3266 | 2454 | 4236 |
| Pre-switch physician reimb. | 2607 | 3258 | 3817 | 2865 |
| Pre-switch physician reimb. v. origin group reimb. per physician | 320 | -207 | -494 | 136 |
| Pre-switch physician reimb. v. destination group reimb. per physician | | 117 | 1363 | -1371 |
| B. Patient characteristics | | | | |
| Mean Age | 75 | 75 | 75 | 75 |
| Share(Male) | .43 | .44 | .44 | .44 |
| Share(White) | .82 | .84 | .85 | .84 |
| Share(Black) | .13 | .11 | .11 | .11 |
| Mean Predicted Mortality | .13 | .13 | .13 | .13 |
| C. Physician characteristics | | | | |
| Share(Male) | .69 | .62 | .64 | .61 |
| Mean Years Experience | 23 | 21 | 21 | 21 |
| Share(Internal Medicine) | 1 | 1 | 1 | 1 |
| D. Group characteristics | | | | |
| Origin Num. Patients | 239 | 254 | 262 | 245 |
| Origin Num. Physicians | 163 | 93 | 83 | 105 |
| Destination Num. Patients | | 303 | 266 | 346 |
| Destination Num. Physicians | | 130 | 120 | 141 |
| $\Delta_{size}$ | | .47 | .46 | .48 |
| Total Physician-Episodes | 34129 | 3242 | 1765 | 1477 |

Column (1) represents overall unadjusted averages/shares for the entire study sample. Column (2) reports averages and shares for the switching physicians, as defined in the text. Column (3) reports statistics for physicians whose destination group is less intense than their origin group, while column (4) represents physicians whose destination group is more intense than their origin group. Reimbursement is abbreviated by reimb., and represents intensity of practice. For switchers, reimbursement measures are calculated in the four quarters prior to the switch quarter, and origin and destination reimbursement measures exclude the index physician. For non-switchers, reimbursement measures are calculated across all quarters, and origin group reimbursement does not exclude the index physician. $\Delta_{pmt}$ is the difference in log intensity between origin and destination groups. $\Delta_{size}$ is the size difference between origin and destination groups, in logs. Years of experience is calculated as 2016 minus the year of graduation from medical school for the sub-sample of physicians who we can match to Physician Compare. Group characteristics are calculated in the four quarters prior to the switch for switchers, and over all quarters for non-switchers. Number of patients in a group is calculated as the average of the distinct number of patients group physicians treated in a hospital in a year-quarter. Number of physicians in a group is calculated as the average of the distinct number of MDs or DOs who treated patients in a hospital in a year-quarter. Total Physician-Episodes reports the total number of distinct physician and switching episodes; each physician can switch multiple times.

Table 2: Pre v. Post-Switch Regressions, $\Delta_{pmt}$, Internists

| *Panel A. Treatment Intensity* | (1) Ln(Pmt) | | (2) Ln(Patients) | | (3) Ln(HCPCS) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $\Delta_{pmt}$*Post Switch | 0.285*** | (0.041) | 0.082*** | (0.016) | 0.141*** | (0.022) | | | | |
| $\Delta_{pmt}$*Qtr=0 | 0.037 | (0.026) | 0.004 | (0.014) | 0.021 | (0.018) | | | | |
| Constant | 7.251*** | (0.001) | 2.184*** | (0.001) | 3.059*** | (0.001) | | | | |
| Adjusted $R^2$ | 0.696 | | 0.774 | | 0.709 | | | | | |
| Mean | 7.250 | | 2.185 | | 3.059 | | | | | |
| *Panel B. Treatment Intensity, Additional Measures* | (4) Ln(Pmt/Pt) | | (5) Ln(Pmt/HCPCS) | | (6) Ln(HCPCS/Pt) | | (7) Ln(Procedures) | | (8) Ln(LOS) | |
| $\Delta_{pmt}$*Post Switch | 0.190*** | (0.030) | 0.128*** | (0.023) | 0.040*** | (0.007) | 0.083*** | (0.016) | 0.004 | (0.005) |
| $\Delta_{pmt}$*Qtr=0 | 0.038* | (0.018) | 0.022 | (0.013) | 0.012 | (0.007) | 0.009 | (0.014) | -0.003 | (0.011) |
| Constant | 5.269*** | (0.001) | 4.310*** | (0.001) | 1.319*** | (0.000) | 1.705*** | (0.001) | 2.046*** | (0.000) |
| Adjusted $R^2$ | 0.674 | | 0.713 | | 0.531 | | 0.708 | | 0.371 | |
| Mean | 5.267 | | 4.309 | | 1.318 | | 1.706 | | 2.046 | |
| *Panel C. Quality of Care* | (9) Share(30-Day Readm) | | (10) Share(30-Day Mort) | | (11) Share(365-Day Mort) | | | | | |
| $\Delta_{pmt}$*Post Switch | 0.0027 | (0.0026) | 0.0034 | (0.0020) | 0.0076* | (0.0033) | | | | |
| $\Delta_{pmt}$*Qtr=0 | -0.0068 | (0.0047) | 0.0037 | (0.0036) | 0.0054 | (0.0062) | | | | |
| Constant | 0.2420*** | (0.0001) | 0.1012*** | (0.0001) | 0.3076*** | (0.0001) | | | | |
| Adjusted $R^2$ | 0.108 | | 0.140 | | 0.179 | | | | | |
| Mean | 0.2423 | | 0.1015 | | 0.3078 | | | | | |
| Observations | 529465 | | 529465 | | 529465 | | 529465 | | 529465 | |

Standard errors in parentheses

This table reports estimated coefficients from Equation 7 for internists. Post Switch is an indicator variable that is equal to 1 for all quarters $\in [1, 10]$. Fixed effects are included for physician-episode and hospital-year-quarter. Standard errors are two-way clustered at the physician and group levels. Omitted category is an indicator for quarters $\in [-10, -1]$. HCPCS is the Healthcare Common Procedural Coding System code recorded as the specific line item in a given claim. "Pt" abbreviates patient

$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

## Table 3: Pre v. Post-Switch Regressions, $\Delta_{size}$, Internists

*Panel A. Treatment Intensity*

| | (1) Ln(Pmt) | | (2) Ln(Patients) | | (3) Ln(HCPCS) | |
|---|---|---|---|---|---|---|
| $\Delta_{size}$*Post Switch | -0.050*** | (0.012) | -0.030*** | (0.008) | -0.042*** | (0.010) |
| $\Delta_{size}$*Qtr=0 | -0.026* | (0.012) | -0.028*** | (0.007) | -0.026** | (0.009) |
| Constant | 7.250*** | (0.001) | 2.184*** | (0.001) | 3.058*** | (0.001) |
| Adjusted $R^2$ | 0.696 | | 0.774 | | 0.709 | |
| Mean | 7.250 | | 2.185 | | 3.059 | |

*Panel B. Treatment Intensity, Additional Measures*

| | (4) Ln(Pmt/Pt) | | (5) Ln(Pmt/HCPCS) | | (6) Ln(HCPCS/Pt) | | (7) Ln(Procedures) | | (8) Ln(LOS) | |
|---|---|---|---|---|---|---|---|---|---|---|
| $\Delta\_size$*Post Switch | -0.015 | (0.008) | -0.005 | (0.006) | -0.007* | (0.003) | -0.026*** | (0.008) | -0.003 | (0.003) |
| $\Delta\_size$*Qtr=0 | 0.005 | (0.007) | 0.000 | (0.005) | 0.004 | (0.003) | -0.022** | (0.007) | 0.003 | (0.005) |
| Constant | 5.268*** | (0.001) | 4.310*** | (0.001) | 1.318*** | (0.000) | 1.704*** | (0.001) | 2.046*** | (0.000) |
| Adjusted $R^2$ | 0.674 | | 0.712 | | 0.531 | | 0.708 | | 0.371 | |
| Mean | 5.267 | | 4.309 | | 1.318 | | 1.706 | | 2.046 | |

*Panel C. Quality of Care*

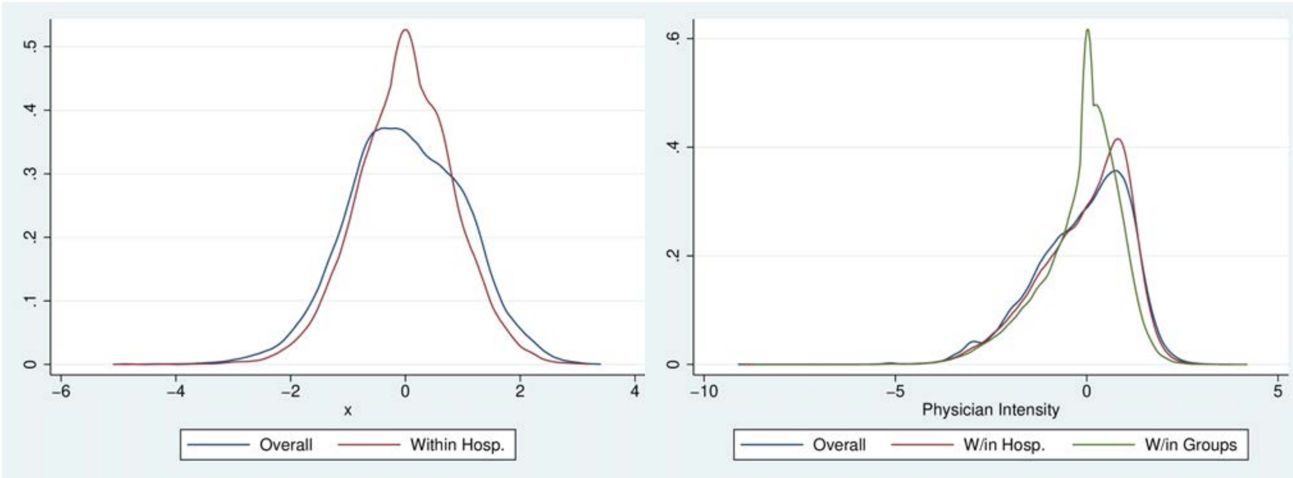
| | (9) Share(30-Day Readm) | | (10) Share(30-Day Mort) | | (11) Share(365-Day Mort) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $\Delta_{size}$*Post Switch | 0.0016 | (0.0013) | 0.0004 | (0.0010) | 0.0003 | (0.0016) | | | | |
| $\Delta_{size}$*Qtr=0 | 0.0004 | (0.0024) | -0.0000 | (0.0019) | 0.0014 | (0.0028) | | | | |
| Constant | 0.2420*** | (0.0001) | 0.1012*** | (0.0001) | 0.3076*** | (0.0002) | | | | |
| Adjusted $R^2$ | 0.108 | | 0.140 | | 0.179 | | | | | |
| Mean | 0.2423 | | 0.1015 | | 0.3078 | | | | | |
| Observations | 529465 | | 529465 | | 529465 | | 529465 | | 529465 | |

Standard errors in parentheses

This table reports estimated coefficients from Equation 7 for internists. Post Switch is an indicator variable that is equal to 1 for all quarters $\in [1, 10]$. Fixed effects are included for physician-episode and hospital-year-quarter. Standard errors are two-way clustered at the physician and group levels. Omitted category is an indicator for quarters $\in [-10, -1]$. HCPCS is the Healthcare Common Procedural Coding System code recorded as the specific line item in a given claim. "Pt" abbreviates patient

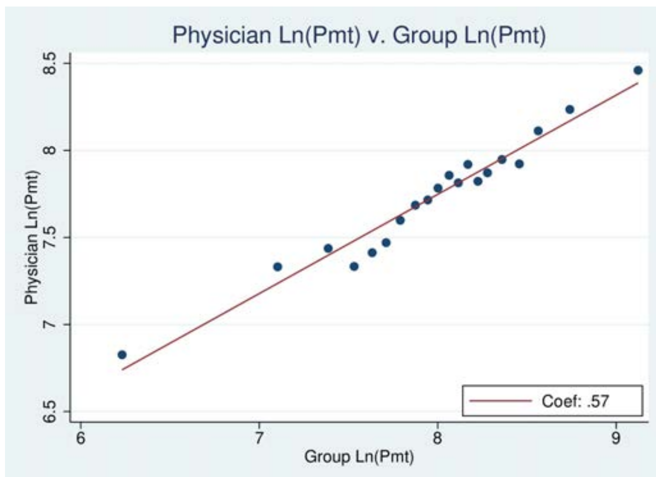
$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

Figure 1: Physician Intensity and Group Intensity, Internists




(a) Group Intensity, Internists

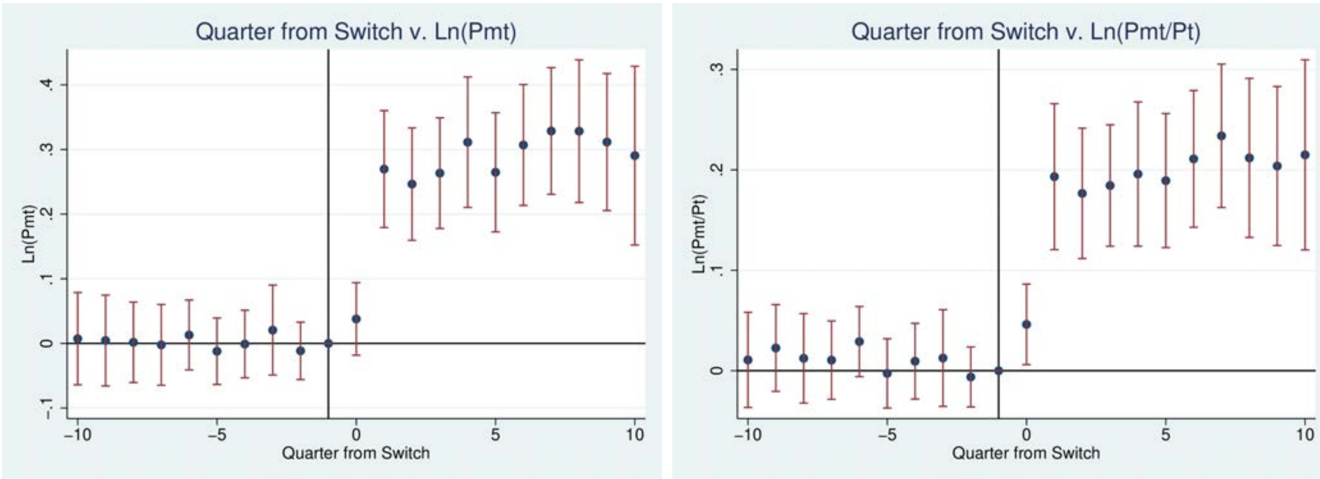
(b) Physician Intensity, Internists



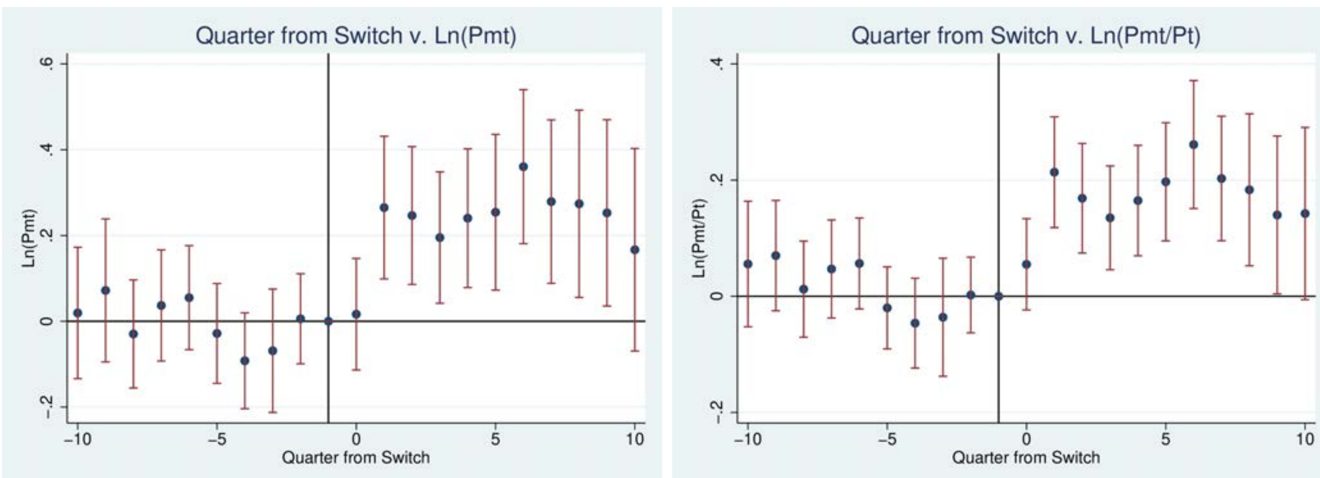

(c) Physician Intensity v. Group Intensity

This figure documents trends in physician and group intensity among internists, as measured by average log reimbursement per physician per quarter, as well as the relationship between the two. Group intensity is calculated as described above, and average physician intensity is calculated across all quarters. Panels (a) and (b) plot the variation in (demeaned) physician and group intensity overall, within hospitals, and within groups (for physicians only), for switchers, non-switchers, and all other out-of-sample physicians associated with in-sample groups. Within-hospital and within-group intensity is demeaned using the hospital- and hospital-group specific averages, respectively. The standard deviation for overall and within-hospital group intensity is 1.01 and 0.85, respectively. The standard deviation for overall, within-hospital, and within-group intensity for physicians is 1.22, 1.18, and 1.03, respectively. Panel (c) plots the relationship between physician intensity and group intensity for switchers. We identify vigintiles of group intensity, and collapse the physician-quarter-level data to averages at these vigintiles, plotted here. The coefficient and standard error are obtained from regression of un-collapsed (i.e. physician-quarter level) physician intensity on group intensity, with no additional controls.

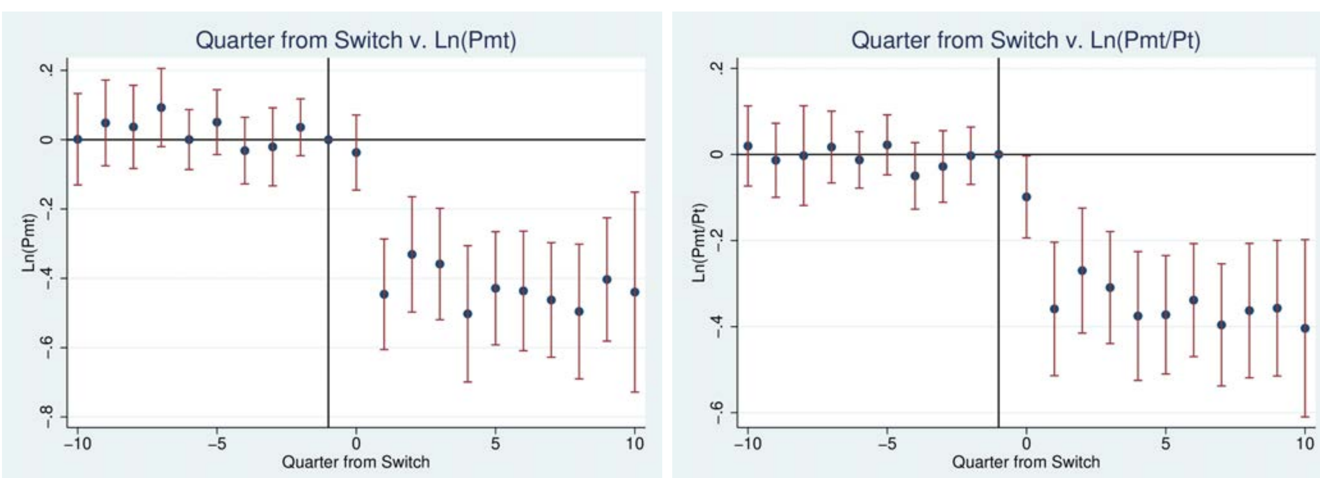
Figure 2: Event Study, Pre v. Post Group Switch, Scaled by $\Delta_{pmt}$, Internists




(a) Pooled $\Delta_{pmt}$



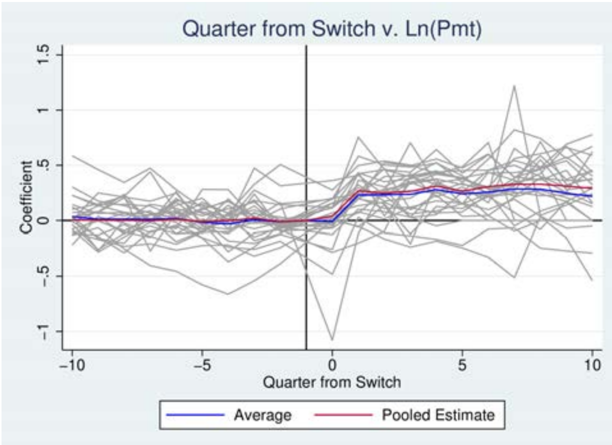

(b) $\Delta_{pmt} > 0$



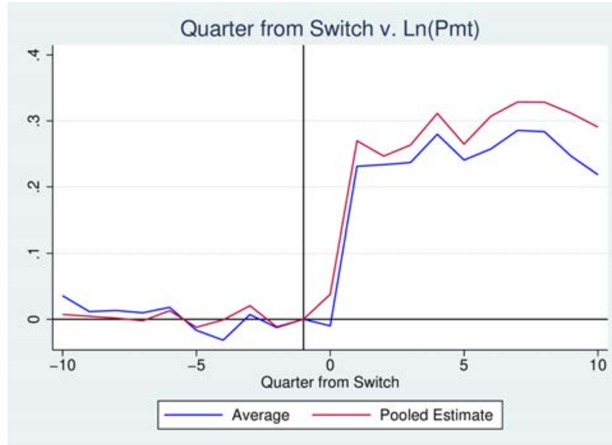

(c) $\Delta_{pmt} < 0$

This figure plots the $\theta_q$'s estimated from Equation 6 for log reimbursement per physician per quarter (left column) and log reimbursement per patient per physician per quarter (right column), scaled by $\Delta_{pmt}$. In panels (b) and (c), we estimate Equation 6 separately for $\Delta_{pmt} > 0$ and $\Delta_{pmt} < 0$, respectively. Included are 95% confidence intervals using standard errors that are two-way clustered at the physician and group level.

Figure 3: Event Study by Switch-Quarter Cohort, Internists



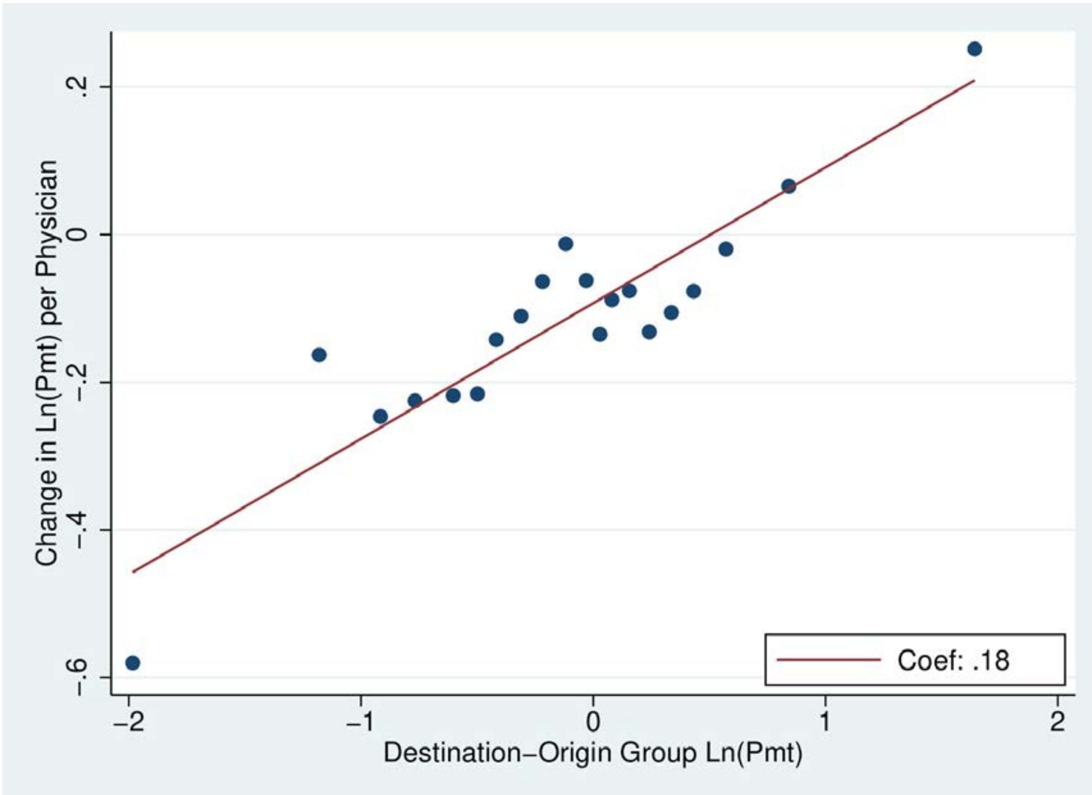

(a) With Switcher Cohorts




(b) Average v. Pooled Estimates

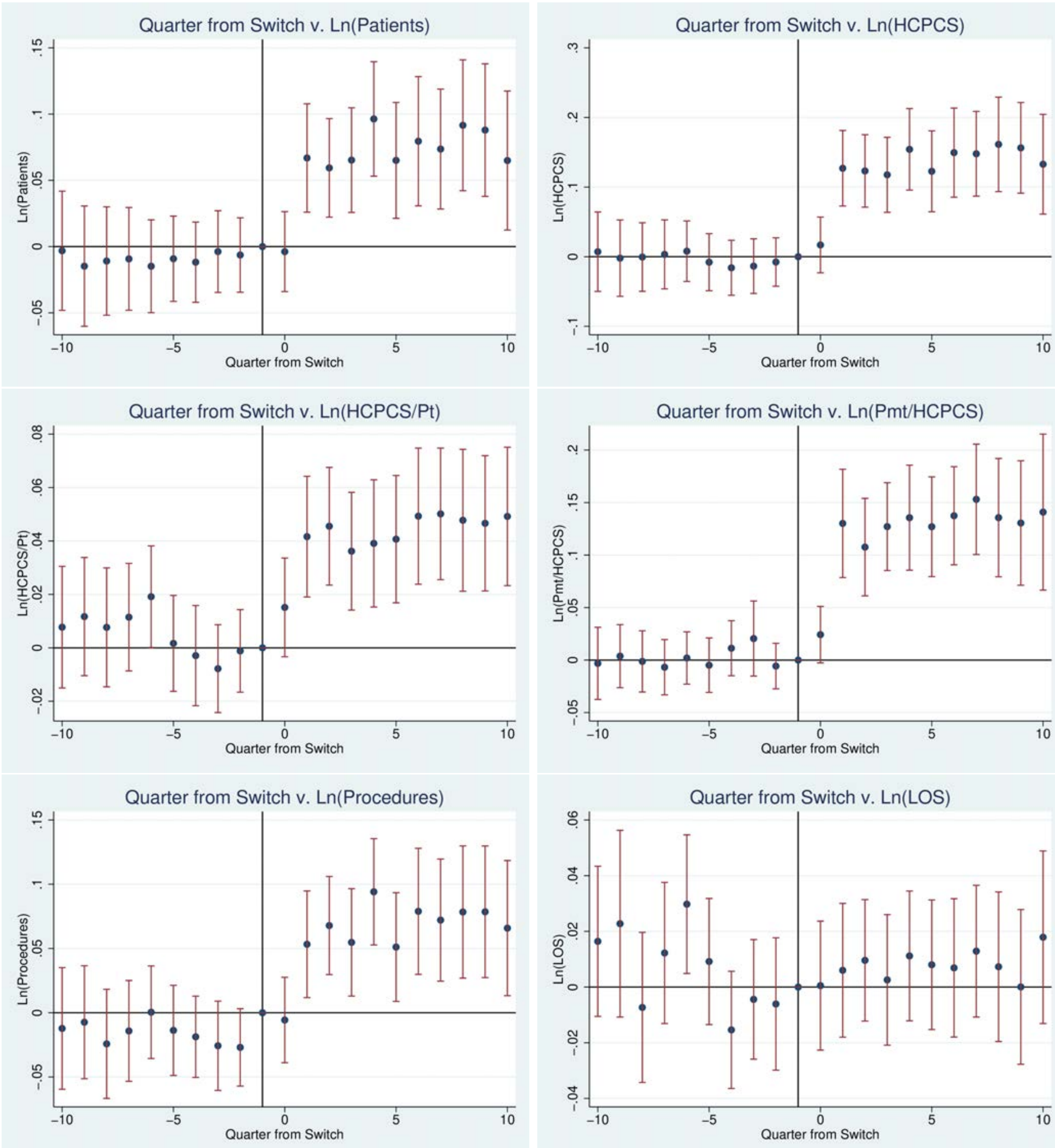
This figure plots the $\theta_q$s obtained from estimating Equation 6 on each cohort of switchers, as defined by quarter of switch. Estimates of difference in intensity relative to switch time (relative to controls) for individual switcher cohorts are plotted in gray in (a). The blue line plots the average of these estimates, and the red line plots the original estimates from running the model on the pooled (all switcher cohort) sample in (a) and (b). 95% confidence intervals are excluded for ease of comparison of the different trend lines.

Figure 4: Changes in Physician-Level Reimbursement per Patient-Quarter v. Destination-Origin Reimbursement per Patient-Quarter Group Differences, by Vigintile, Internists



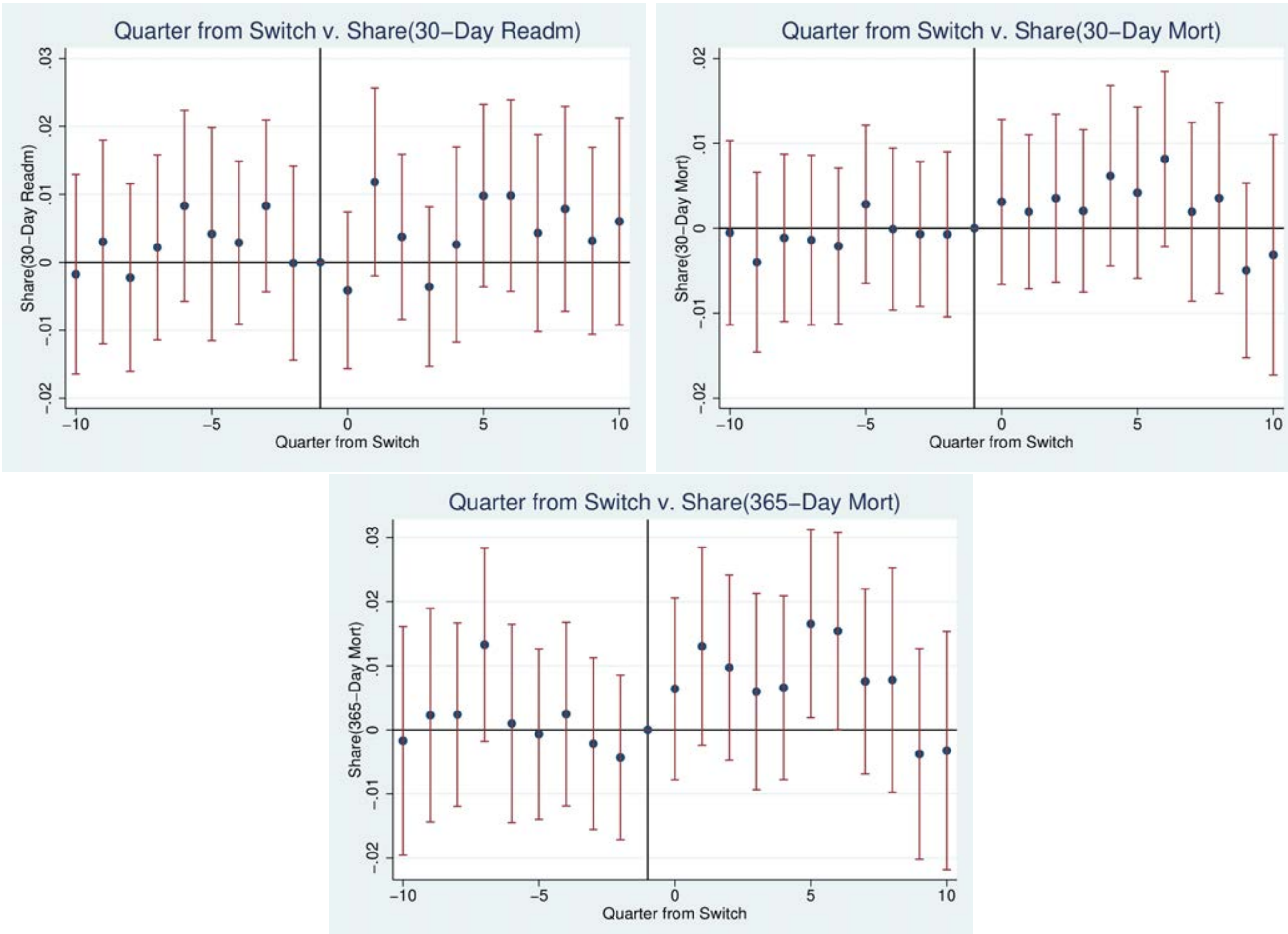

Vigintile plot: This plot is created by identifying the vigintiles of destination-origin group change, and then calculating the average change in outcome at each Vigintile for: i) $\Delta_{pmt}$ (change in reimbursement per patient-quarter across origin and destination groups) (x-axis); ii) $\Delta_y$ (change in reimbursement per patient-quarter at the physician level, calculated in the four quarters around the switch, excluding the switch quarter) (y-axis). The line of best fit is given by a simple OLS regression of the 20 data points associated with $\Delta_y$ on $\Delta_{pmt}$. "Coef" is the slope of the line through these points. The plot focuses on internists.

Figure 5: Event Study, Pre v. Post Group Switch, Scaled by $\Delta_{pmt}$, Internists



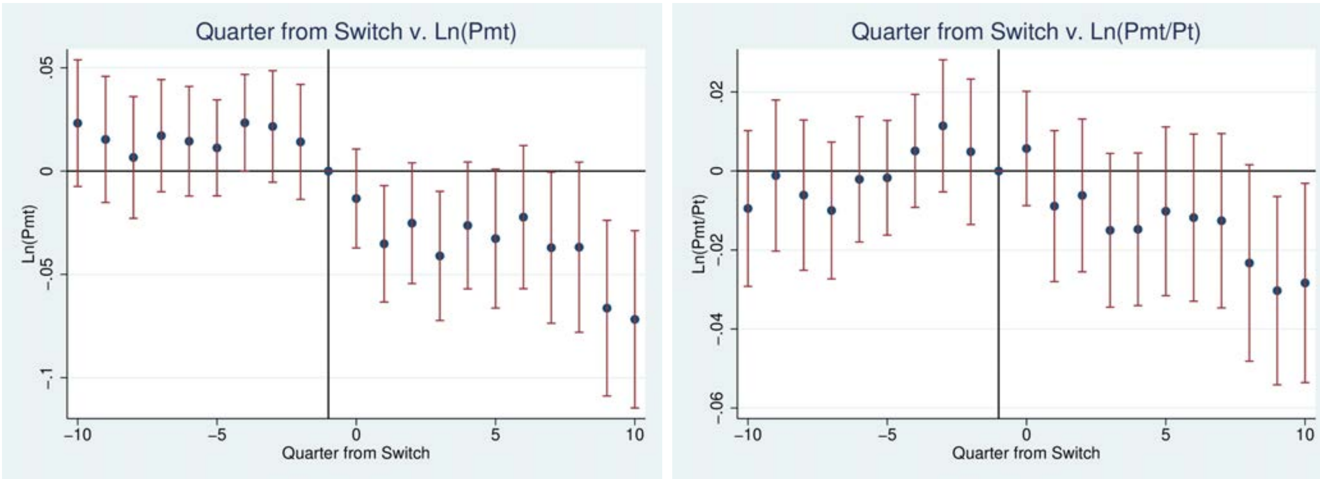

This figure plots the $\theta_q$'s estimated from Equation 6 for additional measures of treatment intensity. The bottom two plots results for measures of treatment intensity that are measured at the hospitalization level (and thus not directly attributable to the index physician. Included are 95% confidence intervals using standard errors that are two-way clustered at the physician and group level.

Figure 6: Quality of Care Event Study, Pre v. Post Group Switch, Scaled by $\Delta_{pmt}$, Internists



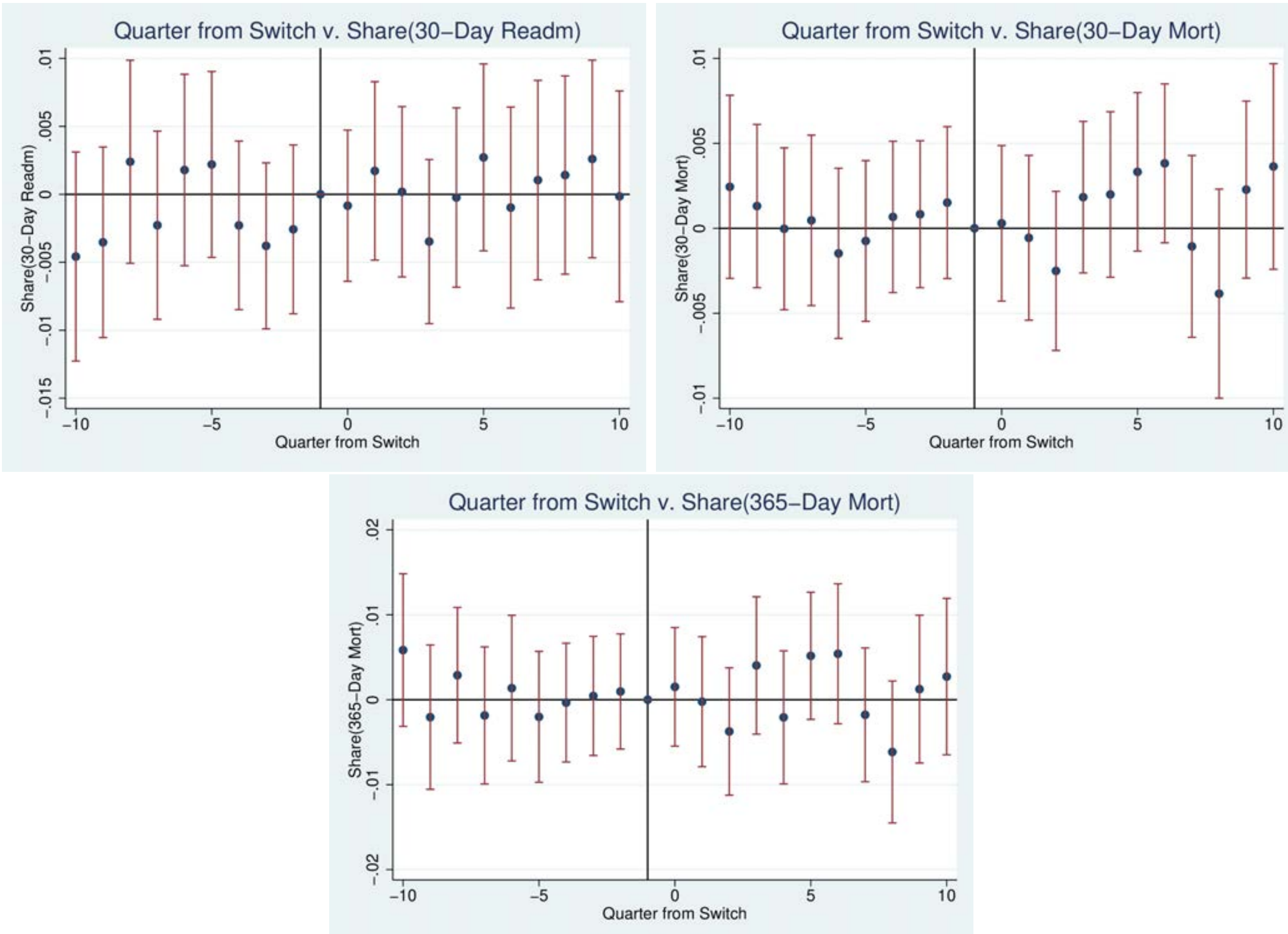

This figure plots the $\theta_q$'s estimated from Equation 6 for measures of quality of care. 30-day readmission rates are calculated as the share of hospitalizations in a given quarter that resulted in a readmission within 30 days of the discharge date. 30- and 365-day mortality rates are calculated as the share of hospitalizations in which the patient died within 30 or 365 days of admission. Included are 95% confidence intervals using standard errors that are two-way clustered at the physician and group level.

Figure 7: Event Study, Pre v. Post Group Switch, Scaled by $\Delta_{size}$, Internists



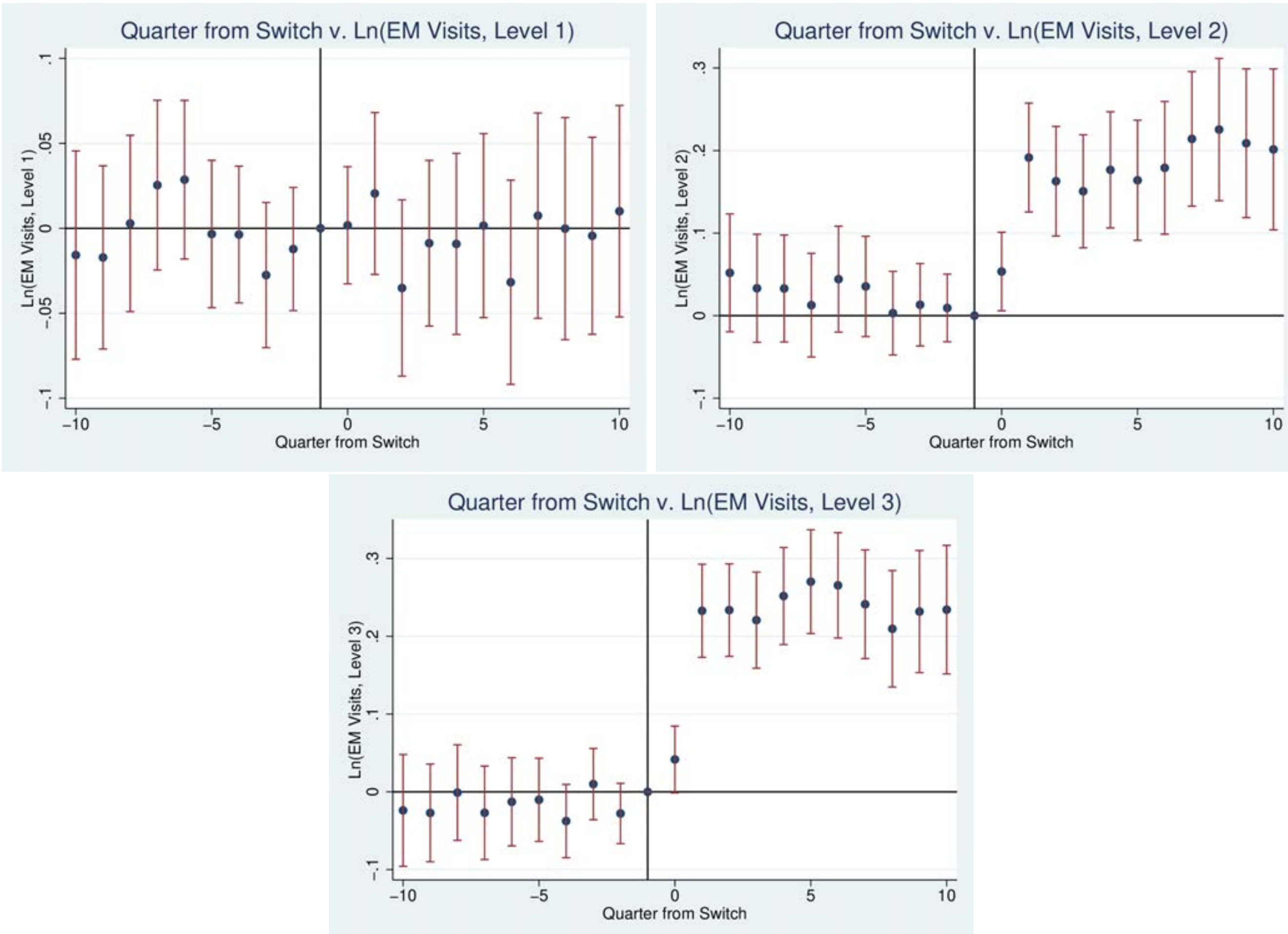

This figure (which is analogous to Figure 2) plots the $\theta_q$'s estimated from Equation 6, substituting $\Delta_{size}$ for $\Delta_{pmt}$, for measures of treatment intensity. Included are 95% confidence intervals using standard errors that are two-way clustered at the physician and group level.

Figure 8: Event Study, Pre v. Post Group Switch, Scaled by $\Delta_{size}$, Internists




This figure (which is analogous to Figure 6) plots the $\theta_q$'s estimated from Equation 6, substituting $\Delta_{size}$ for $\Delta_{pmt}$, for measures of quality of care. Included are 95% confidence intervals using standard errors that are two-way clustered at the physician and group level.

Figure 9: Event Study, Pre v. Post Group Switch, E&M Visits by Level, Internists




This figure (which is analogous to Figure 2)0 plots the $\theta_q$s estimated from Equation 6, scaled by $\Delta_{pmt}$. The outcomes are the log number of E&M visits of a particular level of intensity (levels 1 through 3). Included are 95% confidence intervals using standard errors that are two-way clustered at the physician and group level.