How Do Medicare Payments Influence Physician Practice Structure?

Seth Neller*

November 16, 2020

Abstract

This paper exploits spatial discontinuities in Medicare payment rates to estimate the effect of reimbursements on primary care physicians' choice of organizational structure. I find that a 1 percent increase in Medicare reimbursement leads to a 1.7 to 2.2 percentage point increase in primary care doctors who practice with a small group (defined as 25 providers or fewer). This effect is driven by changes in the tails of the practice size distribution: a 1.8 percentage point increase in physicians who are affiliated with the smallest (1- or 2-provider) practice groups with a corresponding decrease in physicians joining very large practices ($\geq$ 150 providers). I do not, however, detect any evidence of physician sorting or bunching around the boundary in response to differential payment, supporting the underlying assumptions of my regression discontinuity design. Accordingly, my findings suggest that Medicare pricing may be a factor in the trend of consolidation in the physician and clinical services market.

Keywords: Physician payments, Medicare, Practice structure
JEL Codes: H44, H51, I11, I13, I18, L11

*University of Texas at Austin, Department of Economics, 2225 Speedway C3100, Austin, TX 78712. Email: seth.neller@gmail.com. I thank Sam Arenberg, Marika Cabral, Mike Geruso, Sarah Neller, Sam Stripling, Bob Town, and the participants at various UT-Austin Seminars for helpful comments and suggestions. All mistakes are my own.

1 Introduction

In its role as the largest purchaser of clinical services in the United States, the Centers for Medicare and Medicaid Services ("CMS" or "Medicare") yields tremendous influence on physician decisions through its price-setting behavior. The research considering the impact of these prices has largely focused the effects of reimbursement on intensive-margin treatment decisions: i.e. how physicians modify the degree of resource utilization in response to changes in prices. However, an emerging literature has shown the impacts of Medicare payments reach much further, from technology adoption (Finkelstein, 2007; Acemoglu and Finkelstein, 2008; Clemens and Gottlieb, 2014) to private-market price-setting (Clemens and Gottlieb, 2017).

This paper builds on this literature by examining how Medicare policy shapes the way in which physician practices are structured. Specifically, I examine how primary care physicians change their choice of practice structure—as indicated by the size of medical group that they are affiliated with—in response to differential Medicare payment. The structure of physician practices is relevant to both patient health and healthcare costs. Larger multispecialty practice groups may have positive impacts on health outcomes (Epstein et al., 2010; Baker et al., 2019); however, they may also gain bargaining power with commercial insurers, leading to higher prices, premiums, and out-of-pocket costs in the private market (Dunn and Shapiro, 2014; Baker et al., 2014).[1] Given the size of the market for privately financed physician and clinical services ($287 billion in 2016), even limited increases in commercial prices could have large cost ramifications.

To assess the effect of differential reimbursements on these outcomes, this paper exploits spatial discontinuities in Medicare fee-for-service payments within several metropolitan areas across the United States. Medicare has differential payments for providers located in certain "localities," which correspond to select metropolitan areas.[2] Providers practicing outside of these metropolitan areas are assigned a "Rest of State" payment rate, which is typically less generous. Using a boundary regression discontinuity design, I exploit these sharp differences in payment to assess the plausibly exogenous impact of different prices on providers practicing near these borders who serve the same patient populations and face the same input costs.

I find that a 1 percent increase in Medicare reimbursement leads to a 1.7 to 2.2 percentage point ("p.p.") increase in the share of primary care physicians ("PCPs") practicing in smaller medical groups (defined as practices with 25 or fewer providers). This finding is consistent with a theoretical model of physician choice, in which doctors balance financial considerations with non-financial preferences over practice structure. Comparative statics from this model predict that income from more generous reimbursement allows clinicians to adopt practice structures that they find are more suited to their non-financial tastes but that are otherwise less profitable.

To more completely understand how across-the-board changes in Medicare reimbursement af-

[1] Others, such as Scott et al. (2017) and Short and Ho (2019) find no effects or negative effects of vertical integration on quality measures.

[2] These localities often closely follow county boundaries, though they are defined by CMS on the zip-code level. The full list of designated payment localities can be found in Appendix Table A1

fect the nature of medical practices, I examine how differential reimbursement influences the distribution of group sizes. Within my preferred sample, I find that the distribution on the higher-payment side of the boundary stochastically dominates the low-payment distribution. Further analysis underscores that the effects on group size manifest early in the distribution: in response to Medicare reimbursement that is 1 percent more generous, the fraction of primary care doctors practicing in the smallest (1-2 provider) groups increases by 1.8 percentage points. These effects persist through the distribution until they are offset by reductions in PCPs affiliated with groups of 150 or greater. Given that groups of this size are generally comprised of hospital systems or physician groups that are affiliated with a particular hospital system, these findings suggest that higher overall Medicare prices may decrease vertical integration in healthcare markets.

My empirical findings are robust across specifications, bandwidth choices, and sample restrictions. Furthermore, to perform tests of my identifying assumptions, I conduct a variety of bunching tests across different sub-samples and find no evidence that physicians endogenously respond to differential payment by relocating to the side of the border with higher reimbursement. Additionally, I find that observable physician characteristics do not predict whether they receive higher reimbursement, further underscoring that selection is not a driving force of the results.

Finally, I perform a back-of-the-envelope calculation to assess the degree to which these changes in practice structure could result in increases in private-market prices through changes in bargaining power. When combining my estimates with those from Dunn and Shapiro (2014), I find that a 1% *decrease* in Medicare prices may result in a 0.15% to 0.50% *increase* in private prices. Given the relative size of the private market to Medicare, it is possible that a substantial fraction—if not all—of savings from decreased Medicare reimbursement are simply offset by higher costs to participants in the privately financed market.

My findings contribute to the literature that analyzes how Medicare pricing affects market-wide behavior. Acemoglu and Finkelstein (2008) show that a reform in the Medicare Prospective Payment System, which increased the relative cost of labor by moving from full to partial cost reimbursement, increased adoption in new medical technologies in the hospital setting.[3] Additionally, Clemens and Gottlieb (2017) employ a geographic consolidation of payment areas and sharp adjustment to the relative price of surgical versus medical care to investigate how Medicare's prices influence those set by the bargaining processes between providers and private insurers. They find that changes in private prices closely follow those made by Medicare. Finally, the research that is most closely related to this paper is Dranove and Ody (2019), who find that increasing the relative reimbursement for hospital-employed physicians leads to an increase in the share of doctors that vertically integrated with hospital systems (and therefore could participate in these higher reimbursement rates). I build upon their work by finding that, in addition to the directly incentivized

[3]Clemens and Gottlieb (2014) also study technological adoption using a consolidation of geographic adjustments in the Medicare Physician Fee Schedule. They find that increased reimbursement is associated with higher rates of MRI utilization, particularly by non-radiologists. Another paper related to Medicare-induced technology adoption and market structure change—although not directly due to price changes—is Finkelstein (2007), who studies the aggregate effects of the rollout of Medicare.

effects that they estimate, broad pricing pressure can affect market consolidation and integration as well.

This paper also contributes to the body of research studying the impact of prices on physician location. Much of the physician location literature, summarized by Nicholson and Propper (2012), focuses on broad physician location choice in response to financial incentives and the urban-rural divide in provider access and concludes that financial incentives have at most modest effects on location choice. This paper extends this literature by examining whether financial incentives affect location choice within urban areas, as indicated by bunching analyses and other metrics of spatial concentration. Consistent with the literature on broad location choice, I find that financial incentives have no detectable effects.

The remainder of this paper proceeds as follows: Section 2 discusses the background information on Medicare geographic adjustment factors, which provide the identifying variation for my analysis, and the data sources used. Section 3 presents a stylized theoretical model and resulting comparative statics to motivate the expected effect of Medicare payments on practice structure. Section 4 details the identification strategy, econometric specification, and tests of validity for my research design. Section 5 discusses main estimates, robustness checks, and heterogeneity analyses. Section 6 discusses the implications of my findings and concludes.

2 Background and Data

2.1 Medicare Payments and Identifying Variation

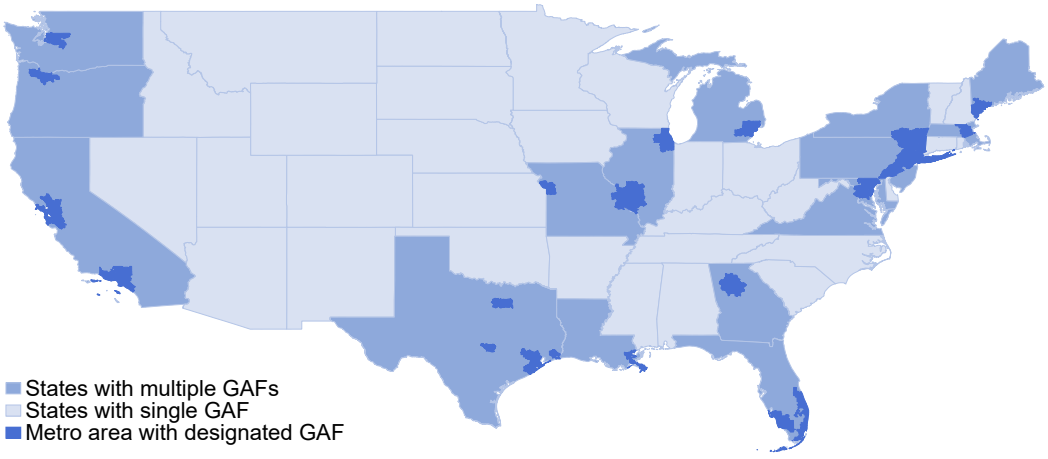
CMS reimburses physician services under Medicare Part B through use of the Physician Fee Schedule. In a given year, providers in payment locality a are reimbursed for service j based on three different components:

$$Payment_{aj} = ConversionFactor \times \mathbf{RVU}_j^{'} \times \mathbf{GAF}_a, \tag{1}$$

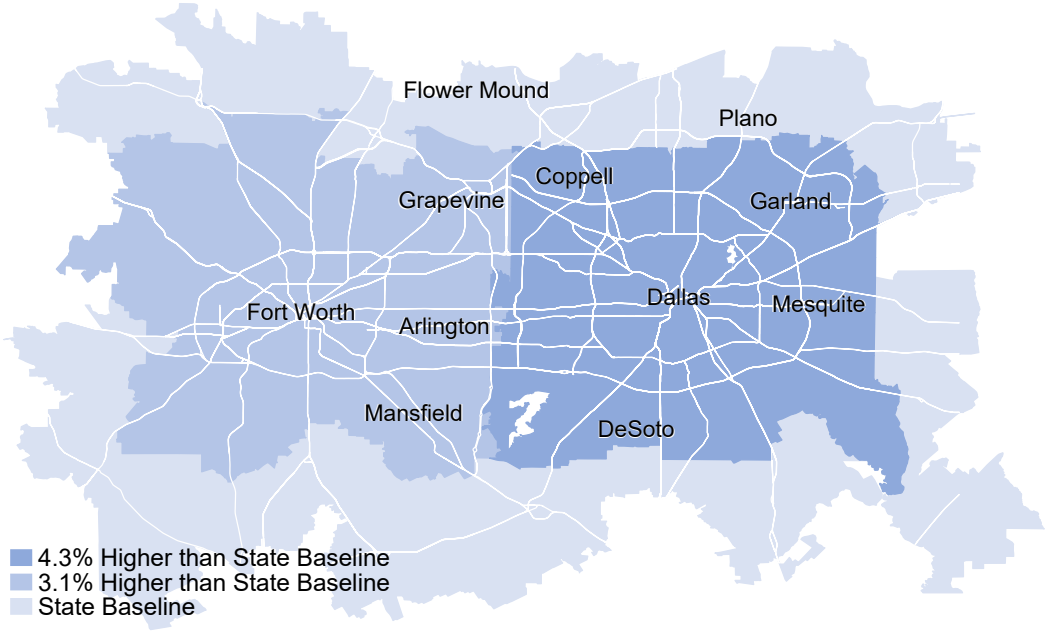
where $ConversionFactor$ is an annually updated baseline for payment and **RVU** represents a vector of three Relative Value Unit ("RVU") components, which are intended to capture the resources required for the labor, practice expense and malpractice premiums, respectively, for a given service (and are therefore a measure of quantity). The **GAF** term represents geographic adjustment factors that account for the fact that labor, practice expense, and malpractice premiums vary in price across various regions.

Medicare-designated metropolitan areas, which have not changed since 1997, receive a GAF that is set based on locality-specific costs, while the remaining areas of the state, including non-designated cities, are assigned a GAF based on costs outside of the designated localities.[4] Figure 1 illustrates the geographic differences in payment areas. The top panel of the figure shows how

[4]The GAF is calculated using physician wages (labor); employee wages, costs for contracted services, rent and supplies (practice expense); as well as liability insurance premiums (malpractice).



(**a**) Payment Localities Across the United States



(**b**) Payment Localities in the Dallas-Fort Worth Area

Notes: The purpose of this figure is to illustrate geographic variation in Medicare payment. Panel A displays differential payments for designated metropolitan areas. There are 17 states that have within-state variation from geographic adjustment by Medicare, which is provided by 37 designated metropolitan areas (indicated by the darkest colors). Light colors indicate states that lack intra-state variation in geographic reimbursement. Panel B illustrates differential geographic adjustments for a large metropolitan area. Primary care providers in the Dallas area and Fort Worth area have payment rates that are 4.3% and 3.1% higher, respectively, than providers of the same health care services in the surrounding suburban areas. Percent differences of geographic adjustment factors are obtained by multiplying each GAF component by the national average of work, practice expense and malpractice RVUs to provide the appropriate weighting of each component. Identifying variation is provided based on the sharp differences in these payments close to the boundaries. Note that the white lines in Panel B represent major roads and highways to provide points of reference.

Source: CMS Physician Fee Schedule.

Figure 1: Identifying Variation: Medicare Payment Locality Boundaries

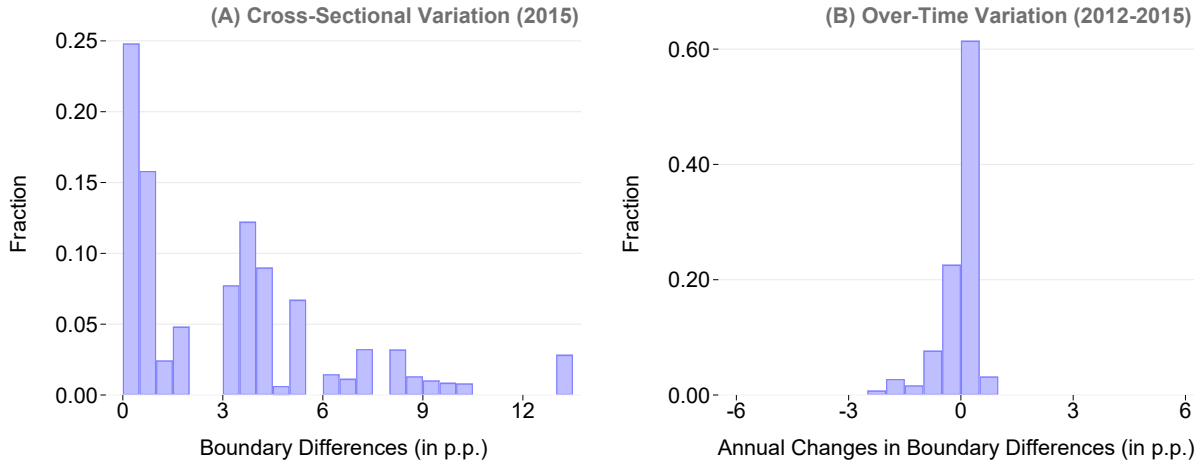


Figure 2: Identifying Variation: Boundary Differences in Medicare Payment

Notes: The purpose of this figure is to illustrate the dispersion of boundary differences in geographic adjustment factors and the relatively stability of those differences over time. Panel A shows the distribution of cross-sectional variation in geographic adjustment across locality borders, where a unit difference is interpreted as a 1% premium in a higher-paying adjacent locality. Panel B demonstrates that these differences are relatively constant over time. Percentage differences of geographic adjustment factors are obtained by multiplying each GAF component by the national average of work, practice expense and malpractice RVUs to provide the appropriate weighting of each component. Density calculations are weighted according to the number of primary care physicians within 4 miles of the boundary.

Source: CMS Physician Fee Schedule and CMS Physician and Other Supplier Public Use File.

different payment localities are dispersed throughout the United States. There are 17 states with multiple geographic adjustment factors and 37 designated metropolitan areas within those states (see Appendix Table A1 for a list of these localities). The bottom panel provides a more detailed view of two localities in the Dallas-Fort Worth Metroplex to illustrate how the resulting payment differences vary over the geography of urban areas. Providers in the Dallas and Fort Worth areas receive more generous reimbursements than providers of the same services in surrounding suburban areas (ex. Plano or Flower Mound). These sharp, localized, *within-state* differences in payments provide the variation used for my analysis. To be clear, I do not utilize across-state differences in geographic adjustment factors in my analysis, since it is likely that many factors (e.g., Medicaid/private insurance environment, licensing requirements, etc.) change across these boundaries.

The magnitude of the differences across geographic boundaries is illustrated by Figure 2, where the unit of analysis is a locality boundary.[5] As evidenced by Panel A, cross-sectional discontinuities between payment areas can be quite large—in some cases over 10%. Furthermore, for providers that face the same costs, as is the case for those close to the boundary for these payments, this additional Medicare revenue translates directly to profit, increasing its potential to impact provider decisions. Meanwhile, these discontinuities are quite stable over time, as detailed in the Panel B.

[5]For instance, the boundary between Dallas and Rest of Texas payment localities is one such locality boundary, as is the border between Dallas and Fort Worth, both of which are displayed in Figure 1(b).

I view this stability as an advantage—Clemens and Gottlieb (2014) demonstrate that physician responses to changes in Medicare prices evolve over several years. Due to the lack of variation in prices over time, I believe the effects that I identify are indicative of long-run behavior.[6]

2.2 Data

The data used for the empirical analysis is primarily gathered from the Centers for Medicare and Medicaid Services. The Medicare Physician and Other Supplier Public Use File supplies provider-level data on addresses, summary information on patients treated, Medicare utilization, and characteristics such as certification (e.g., Medical Doctor, Doctor of Osteopathic Medicine) and primary specialty type. The CMS Physician Compare data set is used to gather additional provider-level information on group size and affiliations, medical school attended, and year graduated.[7] Within the Physician Compare data, providers are considered to be part of a group practice if they have filed a Medicare claim associated with that group's taxpayer identification number within the last twelve months. Because non-physician providers also bill Medicare, the group size measures in this data will differ slightly from those who only measure the number of physicians associated with a given group.

The geographic adjustment factors and relative value units are obtained from the Physician Fee Schedule for reimbursements under Medicare Part B. The American Community Survey ("ACS") five-year (2011-2015) estimates (Ruggles et al., 2019) were used to obtain demographic information on the Census-block-group level, which includes variables on age, educational attainment, insurance coverage, household income, home values, and population. Finally, geographic shape files were obtained from the U.S. Census Bureau. All data used is from 2015 unless otherwise noted.[8] Further information on the data used and sample construction is detailed in Appendix Section B.

To construct the sample of physicians used for analysis, I first geocode all providers in states that have multiple payment localities.[9] Next, using the longitude and latitude for each physician address, I calculated I calculated the straight-line distance between providers and nearby locality borders, keeping all doctors within 10 miles.[10] I exclude from my analysis boundaries where physicians on one side of the boundary do not provide a good counterfactual for those on the other side. This includes boundaries that straddle state lines (because providers on each side face a different regulatory environment, as well as different private and state-level public insurance), as well as those that are defined by rivers (because providers may face different demographic environments). Additionally, I remove borders where there are less than five providers within the

[6] Additionally, boundary discontinuity designs readily lend themselves to the identification of long-term effects, as demonstrated by Ambrus et al. (2020) and Dell (2010), both of which identify long-run impacts from geographic discontinuities.

[7] The medical school rankings in this paper were obtained from Schnell and Currie (2018).

[8] While earlier years are available (specifically 2012-2014), I focus only on the cross-section due to the lack of over-time variation in my treatment, as shown in Figure 2(b), and to avoid confounding my results with after-effects of major legislation, namely the Patient Protection and Affordable Care Act, passed in 2010.

[9] This process was done using Texas A&M GeoServices.

[10] Locality borders are defined by adjacent zip codes that are assigned different geographic adjustment factors.

bandwidth on either side of the border and remove providers that practice in a different locality than either side of the boundary they would otherwise be assigned to.[11]

My final sample is further limited to only include physicians who specialize in primary care; accordingly, all subsequent references to doctors or physicians throughout the remainder of the paper focus on primary care physicians. I focus on primary care physicians for two reasons. First, non-primary care doctors are much more likely than to have multiple practice locations (Xierali, 2018), making treatment assignment based on a single location problematic and likely to introduce measurement error. Second, relative to medical or surgical specialists, doctors in primary care are relatively homogeneous—only two specialties, Family Practice and Internal Medicine, make up roughly 95% of PCPs.[12] This homogeneity readily facilitates analysis within the regression discontinuity framework discussed in Section 4.

Finally, in my preferred sample, I maintain only boundaries that have payment differences of 2 percent or more (I refer to this as my "High-Impact" sample). Unless specifically noted, this sample will be used for all analyses. However, as shown in Section 5.3, my results are robust to including all observations.

3 Theoretical Model

3.1 Setup

While theoretical models of physician behavior have considered how differential prices might induce doctors to change treatment intensity or patient composition between Medicare and privately insured patients—McGuire and Pauly (1991) and Glied and Zivin (2002) are two well-cited examples—there is a lack of theory illustrating how different Medicare payments might induce variation in medical practice size. Accordingly, I present a stylized model of structure choice that contrasts physician profit-seeking behavior with non-financial considerations. However, before doing so, it is useful to understand the considerations that underlie the choices that physicians make when selecting different group sizes.[13] The significant dispersion in medical group size suggests that there are a number of advantages and disadvantages to joining a larger medical group. These advantages and disadvantages are summarized by Casalino et al. (2003), who analyze physician assessments of the advantages of larger group practice—which include bargaining leverage with health plans, economies of scale, and lifestyle factors—as well as the disadvantages—lack of coordination, lack of physician leadership and other problems stemming from group incentives. These issues, along with loss of autonomy over patient care, are echoed in popular literature on

[11] Consider Figure 1(b) for an example of this. A doctor practicing in northwest Coppell could be within a bandwidth-distance of the Fort Worth-Rest of Texas locality border and thus eligible for assignment to that border, even though she practices in the Dallas locality. Accordingly, restrictions are made to ensure that this does not occur.

[12] In contrast, the top two specialist disciplines (Emergency Medicine and Diagnostic Radiology) comprise less than 17% of all specialty providers.

[13] I focus specifically on physicians as decision makers, as they provide the majority of health care services and substantially influence decisions in the clinical setting. This is true even if professional managers make decisions as agents of the clinicians.

the subject (e.g., Lipton 2012).

With these factors in mind, consider a highly simplified model where physicians choose practice structure, s, at the beginning of each period to maximize total utility, u.[14] Total utility is additively separable into utility from the physician's share of profits, $v(\pi(s))$, and utility from non-financial characteristics, $w(s)$. In such a case, the physician solves:

$$\max_s u(s) = v(\pi(s)) - w(s), \tag{2}$$

where both $v(\cdot)$ and $w(\cdot)$ are increasing and concave in s.[15] The assumptions on $v(\cdot)$ are standard, while those on $w(\cdot)$ are motivated by the aforementioned literature—e.g. disutility from coordination and lack of autonomy is increasing in the size of the group practice but exhibits diminishing disutility in larger practices.

Within this model, the physician's profit as a function of group size s is:

$$\pi(s) = p^o(s) \cdot q^o + p^m \cdot q^m - mc(s) \cdot (q^m + q^o), \tag{3}$$

where p^x and q^x denote the price and quantity of patients covered by insurer type $x = \{$Medicare (m), Other private $(o)\}$, and $mc(s) \cdot (q^m + q^o)$ represents a cost function that is linear in quantity and the same for publicly and privately insured patients. The marginal cost function, $mc(s)$, is assumed to be decreasing and convex to capture efficiency costs associated with larger groups. Finally, adapting the conceptual framework from Clemens and Gottlieb (2017), I parameterize private prices, p^o, as:

$$p^o(s) = (1 - \theta(s)) \cdot p^{ins} + \theta(s) \cdot p^m, \tag{4}$$

where $\theta(s) \in [0, 1]$ is the relative bargaining strength of the insurer. The term p^{ins} is the highest price the insurer is willing to pay the physician and p^m is the Medicare price, which acts as the physician's reservation price.[16] Lower values of $\theta(s)$ represent physician groups with substantial bargaining power and are able negotiate prices closer to the insurer's reservation price. Conversely, high values of $\theta(s)$ represent doctors with lower bargaining power and are therefore paid rates closer to the reservation price. To reflect larger groups' ability to elicit higher prices (Dunn and Shapiro, 2014), $\theta(s)$ is assumed to be decreasing and is additionally assumed convex to reflect diminishing bargaining power as s becomes large.

[14] This framework makes two substantial simplifications for the purposes of clarity and tractability. First, the model treats the physician's location as exogenous. As discussed in Section 4.2, this assumption appears to be empirically supported, as there is little-to-no evidence of spatial response to differential payments. Second, quantity could also increase in tandem with higher prices due to physician-induced demand, as demonstrated by Clemens and Gottlieb (2014) or patient composition of Medicare vs. privately insured patients could change. Appendix Section A investigates how incorporating quantity responses affects model outcomes. I find that, under similar assumptions, predictions are largely the same as discussed below. Additionally, the results discussed further in Appendix Section C suggest minimal extensive-margin response to differential prices, supporting the assumption that patient composition is unchanging.

[15] Note that, even though the function $-w(s)$ is convex, it also satisfies the conditions for quasi-concavity based on the strict monotonicity of its first derivative. Accordingly, u is strictly quasi-concave and therefore can achieve a maximum.

[16] Note that, in any equilibrium where bargaining has resulted in an agreement between rational parties, it must be that $p^{ins} \geq p^m$.

3.2 Comparative Statics

To assess the impact of differential Medicare prices on selected group size, consider the first order conditions of the model specified by Equations 2 through 4:

$$\frac{\partial u(s)}{\partial s} = 0 \Leftrightarrow v_\pi \pi_s - w_s = 0 = \phi(s), \tag{5}$$

where subscripts denote derivatives with respect to the subscripted argument. Then, totally differentiating to obtain the comparative static of Medicare prices, p^m, on group size, s, yields:

$$\frac{ds}{dp^m} = -\frac{\partial \phi(s)/\partial p^m}{\partial \phi(s)/\partial s} = -\frac{v_{\pi\pi}\pi_s \pi_{p^m} + v_\pi \pi_{sp^m}}{u_{ss}}. \tag{6}$$

Utilizing the properties of the model specified in Section 3.1, the sign of Equation 6 is:

$$\frac{ds}{dp^m} < 0. \tag{7}$$

Thus, the model predicts that group size responds negatively to higher in Medicare prices, as the income effect of higher prices affords physicians the luxury of choosing group sizes that more closely align with their non-financial preferences. This conclusion holds even when price or marginal costs are assumed to be unresponsive to group size (i.e. $\theta_s = 0$ or $mc_s = 0$). I next consider whether this conclusion is empirically true in the following sections.

4 Empirical Framework

4.1 Identification and Econometric Model

To identify the effects of Medicare payments on practice structure, I estimate the following model for all providers within the neighborhood of a locality border:

$$Y_{ip} = \delta \cdot High_p + \phi_{b(p)} + f(Z_p) + h(X_{ip}) + \varepsilon_{ip}, \tag{8}$$

where i denotes a primary care physician practicing in location p. The main outcome of interest, Y_{ip}, is a variable indicating whether a PCP is affiliated with a small practice group, defined as a medical group practice with 25 or fewer providers. The threshold of 25 providers was chosen because that is the group size generally used by CMS to distinguish small and large practice groups.[17]

[17] See, for instance, Centers for Medicare and Medicaid Services (2015). This measure was chosen as my primary measure, rather than a measure of average group size selected, because moderate shifts in the *distribution* of group sizes can have very large effects on the *average* group size that are difficult to interpret and may seem implausible at first glance. To be concrete: in the context I study, if 5% of primary care physicians in a market moved from the 75th percentile of group size (roughly 250 providers) to the 25th percentile of group size (roughly 2 providers), this would result in a reduction of ∼0.50 log points (or a ∼40% decrease) in the average log group size chosen by physicians in that market. Nonetheless, I also consider the effects of this payment discontinuity on average (log) group size and find that the effects are consistent with the other results presented in this paper. See Appendix Section C for more detail.

The variable $High$ is a binary variable equal to one if a provider is on the side of the border with higher payment rates. The coefficient on $High$, δ, is the parameter of interest. To account for locality-specific characteristics of providers, I also include border fixed effects, $\phi_{b(p)}$. In certain specifications, $\phi_{b(p)}$ represents cluster-by-border fixed effects, wherein nearby physicians assigned to the same border are grouped together into clusters (using k-means clustering) in order to better control for spatial characteristics of the localities that I study. Next, $f(Z_{ip})$ represents a polynomial in distance from the border that varies on each side of the cutoff. In my preferred specification, I utilize a first-order polynomial. Finally, $h(X_{ip})$ is comprised of a flexible set of local, patient-population, and provider controls.[18]

The identifying variation from my analysis comes from sharp changes in geographic adjustment factors over space. I assign GAFs to providers based on their practice location and specialty. Specifically,

$$GAF_{ip} = \sum_{\kappa \in work,pe,mp} gaf_{\kappa(p)} \cdot rvu\_share_{\kappa(i)}, \tag{9}$$

$gaf_{\kappa(p)}$ is the geographic adjustment factor component (work, professional expense, or malpractice premiums) that varies based on physician zip code and $rvu\_share_{\kappa(i)}$ is the share of relative value units in each component for physician i's specialty, calculated at the national level.[19] The GAF is constructed using national RVU shares—-rather than individual shares—-to avoid any endogenous change in service mix in response to differential payment.

The interpretation of the coefficient of interest, δ, in Equation 8 is the change in the outcome in response to higher payment. To facilitate interpretation as the effect of a 1 percent change in reimbursement, I also estimate a regression discontinuity-instrumental variables ("RD-IV") design as follows:

$$Y_{ip} = \delta \cdot \widehat{GAF}_{ip} + \phi_{b(p)} + f(Z_p) + h(X_{ip}) + \varepsilon_{ip}, \tag{10}$$

$$GAF_{ip} = \theta \cdot High_p + \phi_{b(p)} + f(Z_p) + h(X_{ip}) + \varepsilon_{ip}. \tag{11}$$

The coefficient θ in the first stage (Equation 11) describes the average differential in reimbursement between the high- and low-payment sides of the borders in my sample. In my main analysis, the coefficient of interest, δ, is equal to the percentage point change in the number of PCPs practicing in small groups in response to a 1 percent increase in Medicare generosity. For nota-

[18] Local controls include percentage of population by (a) age bins, (b) educational attainment bins, and (c) insurance coverage categories; median household income, median real-estate value, and density controls. Block group characteristics are assigned to individual physicians by calculating an inverse-distance-weighted average of characteristics within 2 miles of a physician's location, where distance is calculated from the physician location to block-group population centroid. Patient population controls include the average age and hierarchical condition category score for a physician's patient population. Finally, physician characteristics include quantiles of medical school rank and experience, as well as controls for gender.

[19] To provide a concrete example: if 53%, 43%, and 3% of the Family Practice specialty's total RVUs come from work, professional expense, and malpractice RVUs, respectively, then these percentages will be used as weights for the locality-specific work, professional expense, and malpractice GAFs to arrive at the final geographic adjustment factor for that specialty.

tional convenience (and to differentiate from estimates from other equations), I refer to this coefficient as "δ_{RD-IV}" throughout the paper. This estimate could alternatively be recovered by separately estimating the first-stage regression (Equation 11) and the reduced-form equation (Equation 8) and subsequently dividing the reduced-form estimate (δ_{RF}) by the first-stage estimate ($\delta_{RD-IV} = \delta_{RF}/\theta$).

The estimation of my reduced-form analysis (Equation 8) is performed using ordinary least squares, while RD-IV analyses are estimated using two-stage least squares. All analyses utilize a rectangular kernel in distance from the border and standard errors clustered at the census tract-level to account for spatial correlation. Alternatives to both kernel and clustering choices are considered in Appendix Section C, with highly similar results.

The selected bandwidth is based on the travel distance to obtain medical services. Specifically, I set bandwidth equal to the median travel distance in the core-based statistical areas that contain my borders of interest, as determined by the National Household Transportation Survey ("NHTS"). The resulting bandwidth is equal to 4 miles. Robustness to alternate bandwidth choices are considered in Section 5.2.

4.2 Validity of Regression Discontinuity Design

The primary identifying assumption underlying my regression discontinuity design is that the unobservable determinants of practice structure evolve smoothly across the boundary (i.e. that they are continuous with respect to the running variable). However, in the context studied in this paper, there are clear financial incentives for physicians to migrate to higher payment areas. Accordingly, a threat to my research design is manipulation of the running variable, which would be evidenced by significant high-payment side bunching immediately after the boundary, as providers relocate their practices to more profitable locations.

I consider such behavior in Figure 3. Panel A displays the results of a McCrary (2008) bunching test, which shows no evidence of manipulation around the boundary. However, the figure does present a distinct bi-modal pattern, with peaks on either side of the boundary (with a higher-density mass on the high-payment side), which is largely due to changes in population density. Accordingly, Panel B presents a similar analysis to Panel A using the total number of primary care physicians per thousand individuals. Within the figure, each dot represents physician-to-population ratios in one-tenth-of-a-mile increments, along with local linear regression lines and associated 90 percent confidence intervals fit to either side of the cutoff. This figure displays no detectable discontinuity at the cutoff. Notably, there is a statistically insignificant *decrease*—rather than the hypothesized increase—in the physician-to-population ratio on the high-payment side of the boundary.

A further concern regarding manipulation of the running variable is that, due to spatial frictions, physicians might relocate several miles away from the cutoff, rather than immediately to the high-payment side. In this case, it would be possible that there would be no evidence of bunching at the cutoff, but there would still be mass displaced from the low-payment side to the high-payment

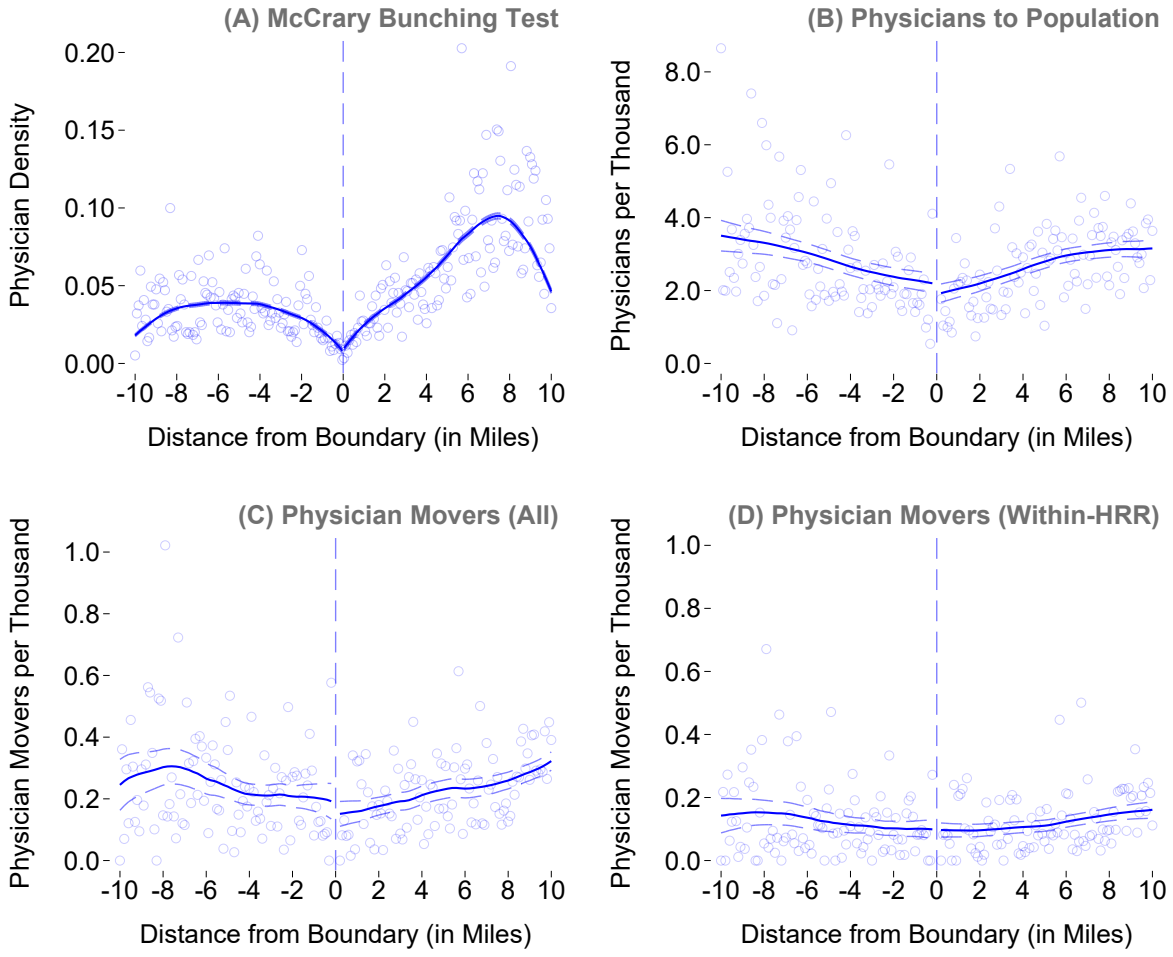


Figure 3: Validation of Research Design: Spatial Provider Density

Notes: The purpose of this figure is to illustrate the absence of bunching around locality boundaries. Panel A performs a McCrary (2008) analysis to assess the degree of bunching, if any, around the cutoff. Note that negative values indicate distances in the relatively lower payment areas, and positive values indicate distances in the higher payment areas. Panel B shows the spatial distribution of primary care physicians per 1,000 individuals by their distance to the nearest treatment boundary. Panels C and D examine physicians who moved to their current location during the years 2012-15, with Panel C examining all such movers and Panel D displaying results for movers who stayed within the same Hospital Referral Region ("HRR"). All local linear regression lines are presented with their 90 percent confidence intervals. For presentation purposes, observations are winsorized at the 99th percentile. See also Appendix Figure A1 for other manipulation tests. This analysis utilizes the High-Impact Sample discussed in Section 2.2.

Source: Author calculations using CMS Physician Fee Schedule and CMS Physician and Other Supplier Public Use File.

side. If this were the case, then Figure 3B would demonstrate a consistently higher mass of primary care physicians per thousand individuals on the high-payment sides of the boundary, which I do not see. To further extend the analysis, Panels C and D of Figure 3 examine the location choices of PCPs who moved within the last three years, on the basis that physician-movers may be subject to fewer optimization frictions as non-moving physicians. Panel D further limits these movers to those that have moved within hospital referral regions ("HRRs") and thus may be more informed

about local differences in reimbursement than other movers. Neither of these "mover" analyses provide any evidence to suggest that primary care physicians are sorting around these payment boundaries, even when optimization and informational frictions are reduced. Finally, Appendix Figure A1 considers additional tests of bunching by re-performing two different variants of the analysis in Figure 3B: (i) using the log (rather than level) of the physician-to-population ratio and (ii) using the number of practice *locations* (rather than primary care physicians). The results of these analyses closely mirror those in Figure 3B, further supporting a lack of sorting around the boundary.

It is worth considering why substantial bunching is not observed at the cutoff, given the impact of differential payments on other aspects of provider practice. There are several possible reasons for this. First, geographic location plays a significant role in patient demand for health care services (e.g., Dunn and Shapiro, 2014; Gowrisankaran et al., 2015; Kessler and McClellan, 2000). Accordingly, relocation of a practice could result in significant patient turnover and thus could be very costly. A second and closely related factor is the impact of spatial proximity on competition between medical practices. To the degree that patients choose providers based on distance, providers will be incentivized to avoid increases in competition—and the resulting patient acquisition costs and/or loss of volume—that would result from bunching. Finally, while the impact of these payments on physicians' profit is substantial, the actual borders that determine these price differences may not be salient for physicians. This is particularly true for physicians (and groups) who practice in a single locality, as they may not have a readily available counterfactual of the differential payment they would receive if practicing elsewhere. Additionally, while physician groups with multi-locality practices may be more aware of geographic payment differences, they are disincentivized to spatially adjust, because doing so might lead to competition among physicians in the same practice.

While I do not detect any manipulation of the running variable, it is theoretically possible that physician sorting across the boundary is occurring in a way that is not detected by my bunching analysis. While I cannot test for sorting along unobservable dimensions, I can evaluate whether discontinuous changes in observable physician characteristics exist. To evaluate discontinuities in observable characteristics, I apply the least absolute shrinkage and selection operator ("LASSO") method to predict (i) receipt of treatment and (ii) choice of small group practice. Employing a LASSO design in this context allows me to utilize a wide range of physician characteristics and their interactions to best predict my outcomes of interest and examine their evolution over the cutoff.[20] Insofar as sorting on either side of the boundary occurs, I expect the LASSO-predicted outcomes to also display discontinuous changes at the cutoff.

Figure 4 displays the results of this analysis, where Panel A shows the results when the LASSO-predicted GAF is the outcome variable in Equation 8 and Panel B demonstrates the estimated effect when the outcome is the LASSO-predicted fraction of primary care physicians in a small practice

[20]Physician characteristics include indicator variables for specialty type, gender, credential, medical school ranking quantile, experience quantile, and quantiles for a wide array for patient characteristics. Interacting these indicators to allow a highly flexible specification equation yields 1,931 covariates for the LASSO procedure to select from.

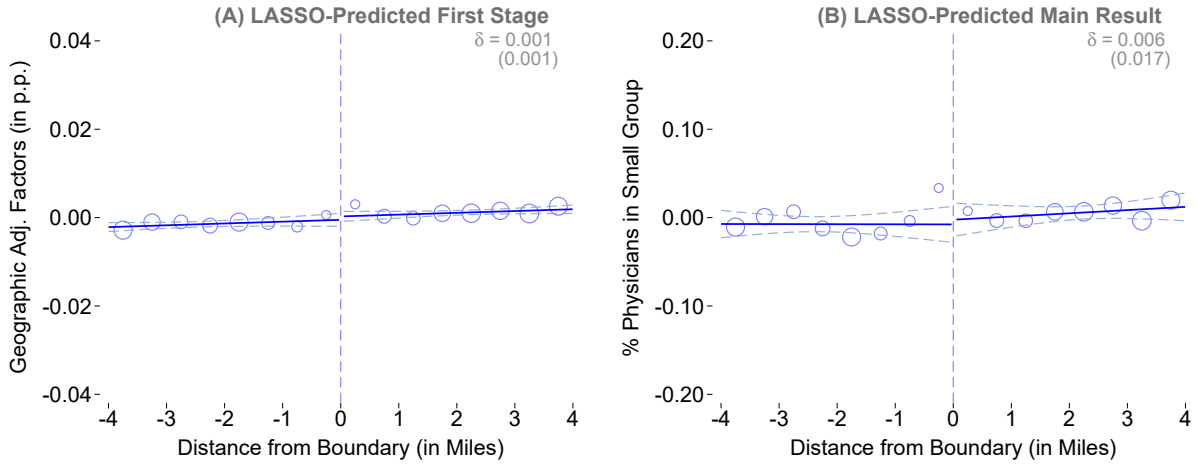


Figure 4: Validation of Research Design: Smoothness of LASSO-Predicted Outcomes

Notes: The purpose of this figure is to illustrate the smoothness of covariate-predicted outcomes through the cutoff. As further discussed in the text, each panel represents a LASSO-generated prediction using physician characteristics (and their interactions) to establish that observable characteristics are not predictive of either the treatment (Panel A) or main outcome (Panel B). Within the panels, each point represents the means of the outcome variable (displayed on the y-axis) in 0.5-mile increments, with the size of the point varying by the number of observations included. All means presented have been regression-adjusted to net out locality-border fixed effects. Each side of the cutoff—where negative distances represent low-payment sides and positive distances are associated with high-payment sides—are fit with linear regression lines and associated 90% confidence intervals. Estimated coefficients from Equation 8 along with associated standard errors are presented in the upper-left corner. This analysis utilizes the High-Impact Sample discussed in Section 2.2.

Source: Author calculations using CMS Physician Fee Schedule, CMS Physician Compare Data, CMS Physician and Other Supplier Public Use File.

group. Within the panels, each point represents the means of the outcome in 0.5-mile increments, with the size of the point varying by the number of observations included. All means presented have been regression-adjusted to net out locality-border fixed effects. Each side of the cutoff—where negative distances represent low-payment sides and positive distances are associated with high-payment sides—are fit with linear regression lines and associated 90% confidence intervals. Estimated coefficients from Equation 8 along with associated standard errors are presented in the upper-left corner. (All regression discontinuity plots displayed in the remainder of this paper are formatted in this way.)

As displayed in the figure, there is no evidence of sorting across the cutoff. Not only do the point estimates fail to achieve any traditional level of statistical significance, but they are also economically small: a 0.1 percent difference in the predicted geographic adjustment factor and a 0.6 percentage point difference in predicted small group participation, both of which are 1-2 orders of magnitude smaller than the results discussed in the next section. Taken in total, the analyses suggest no physician sorting that would bias my results. As an additional test of robustness, I further examine the sensitivity of my estimates to inclusion of provider and block-group controls in Section 5 and find that my results are not sensitive to inclusion of covariates.[21]

[21] As further discussed in Appendix Section C and displayed in Appendix Figure A2, I also examine how individual

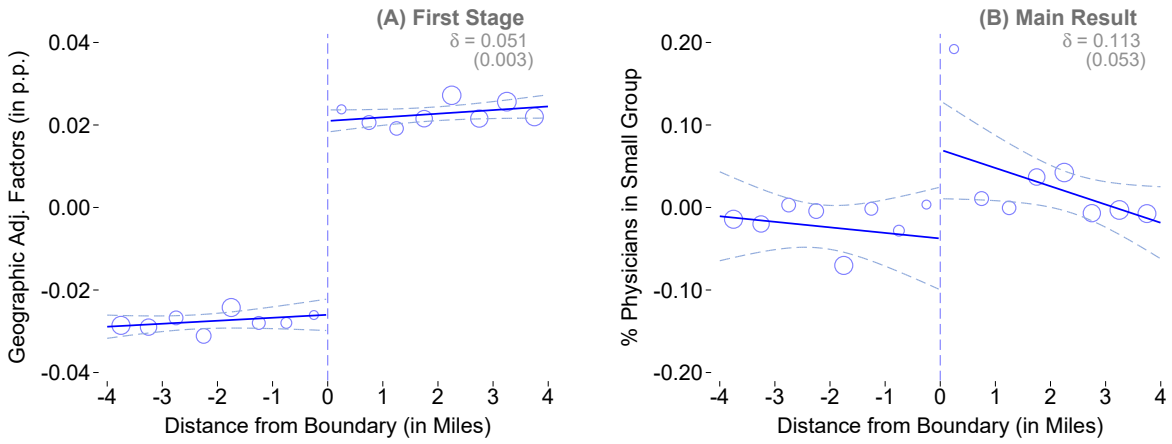


Figure 5: Main Results: Discontinuities at the Boundary

Notes: The purpose of this figure is to display the main results of my analysis: the first-stage impact of the border discontinuity (Panel A) and the second-stage impact on the percentage of primary care physicians practicing in small groups (i.e. those with 25 or fewer providers, displayed in Panel B). The coefficients and associated standard errors displayed in each panel are estimated using Equation 8. All means presented have been regression-adjusted to net out locality-border fixed effects. Each side of the cutoff—where negative distances represent low-payment sides and positive distances are associated with high-payment sides—are fit with linear regression lines and associated 90% confidence intervals. Estimated coefficients from Equation 8 along with associated standard errors are presented in the upper-left corner. Finally, for context, note that the pre-cutoff raw means corresponding to Panels A and B are 1.00 and 0.44, respectively. This analysis utilizes the High-Impact Sample discussed in Section 2.2.

Source: Author calculations using CMS Physician Fee Schedule, CMS Physician Compare Data, and CMS Physician and Other Supplier Public Use File.

5 Results

5.1 Main results

The central results of my analysis are graphically displayed in Figure 5. Panel A demonstrates the impact of the border discontinuity on physician reimbursements: among borders included in my High-Impact Sample, the discontinuity results in a 5.1 percent increase in reimbursement on average. Panel B displays the effect of this discontinuous increase in payments on physician practice structure. Consistent with the comparative statistics of the model described in Section 3, I find that primary care physicians organize into smaller groups in response to more generous payment. Specifically, I estimate that there is an 11.3 percentage point increase in primary care physicians that practice in small groups—defined as groups with 25 or fewer providers—on higher-payment sides of the boundary. The estimates in this figure are mirrored in the first column of Table 1, Panels A and B, respectively. An unusual feature of the regression discontinuity plot in Figure 5, Panel B worth noting is that the point immediately to the right of the cutoff has a substantially higher fraction of PCPs in small group practices than would be expected based on the regression line. To

characteristics vary across the cutoff, both on the physician- and block-group-level. These results also do not suggest any physician sorting across the boundary.

Table 1: Main Results

Outcome	**(1)**		**(2)**		**(3)**		**(4)**	
			Panel A: First Stage					
Geographic Adjustment Factors	0.051	***	0.051	***	0.051	***	0.049	***
	(0.003)		(0.003)		(0.002)		(0.002)	
			Panel B: Reduced Form					
% of Physicians in Small Groups	0.113	**	0.093	*	0.098	*	0.101	**
	(0.053)		(0.051)		(0.052)		(0.051)	
			Panel C: Instrumental Variables					
% of Physicians in Small Groups	2.208	**	1.819	*	1.963	*	2.055	**
	(1.052)		(1.009)		(1.026)		(1.046)	
Controls:								
Distance controls	X		X		X		X	
Border FE	X		X		X			
Provider characteristics			X		X			
Local controls					X			
Cluster FE							X	
Dep. var. mean	0.456		0.456		0.456		0.456	
Observations	5,065		5,065		5,065		5,065	
First stage F-statistic	399		396		505		541	
Bandwidth (in miles)	4		4		4		4	

Notes: This table reports the results of the analysis for different specifications. Panel A describes the estimation of the first stage (i.e. when GAF_{ip} is an outcome) using Equation 8. Panel B estimates the effect of the discontinuity on physician practice structure using Equation 8, where the coefficients are interpreted as the reduced-form effect of crossing the payment boundary. Panel C displays coefficients when estimating the RD-IV strategy (Equations 10 and 11), wherein coefficients are interpreted as the effect (in p.p.) of a 1 percent increase in reimbursement. Columns 1 through 4 present the baseline specification and specifications with differing controls. All specifications utilize a rectangular kernel in distance from the boundary and are clustered at the Census-tract level. This analysis utilizes the High-Impact Sample discussed in Section 2.2. Significance at 10%, 5%, and 1% levels is represented by *, **, and ***, respectively.

Source: Author calculations using CMS Physician Fee Schedule, CMS Physician Compare Data, CMS Physician and Other Supplier Public Use File, Census Block Group 5-year ACS Estimates (Ruggles et al., 2019).

alleviate concerns that this outlier point is driving the results of my analysis, I perform a "donut-hole" test in Section 5.2, wherein I re-perform my analyses after omitting observations within a specified distance of the cutoff. I find that omission of this unusual point does not meaningfully change my results.

To facilitate the interpretation of the effects that I find in my reduced-form analysis, I estimate the RD-IV specification (Equations 10 and 11) in Panel C of Table 1. These estimates imply that a 1 percent increase in the reimbursement rate is associated with a 2.2 p.p. increase in the fraction of primary care doctors affiliated with small practice groups. As illustrated in Columns 2 and 3

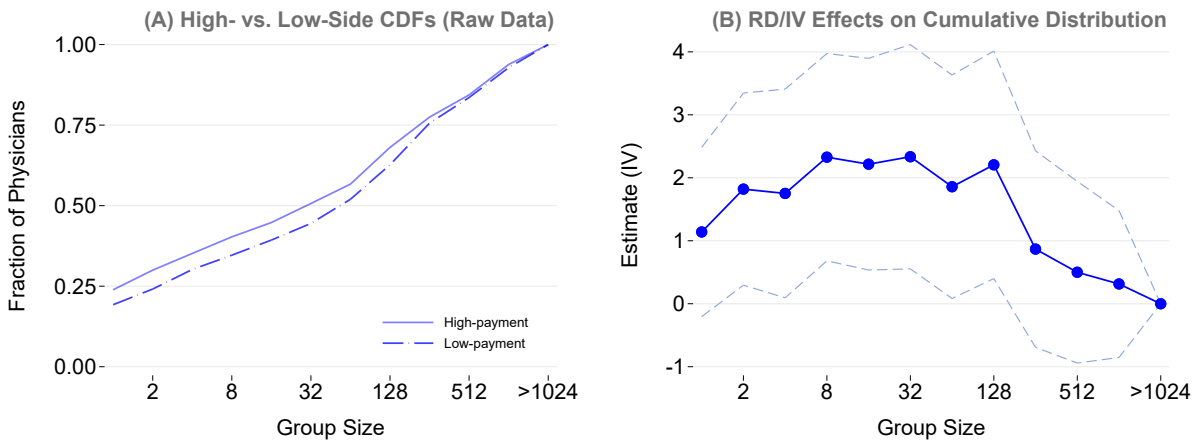


Figure 6: Distributional Effects on Group Size

Notes: The purpose of this figure is to display the distributional effects of differential Medicare reimbursement on group size. Panel A separately displays raw cumulative distribution functions of physicians by practice group size for high- and low-reimbursement sides of the border and within the bandwidth. Panel B displays point estimates—and associated 90% confidence intervals—for the RD-IV specification in separate regressions, where the outcome $Y_{ip} = 1(GroupSize \leq g)$ is generated for each value of group size g on the x-axis. This analysis utilizes the High-Impact Sample discussed in Section 2.2.

Source: Author calculations using CMS Physician Fee Schedule, CMS Physician Compare Data, and CMS Physician and Other Supplier Public Use File.

of Table 1, these results are stable when control variables are added. Furthermore, when I replace my border fixed effects with more spatially fine-grained cluster-by-border fixed effects (Column 4), I obtain similar results.[22] As noted in Section 4, all specifications utilize a rectangular kernel in distance from the boundary and are clustered at the Census-tract level. Because the results are highly similar across all specifications, I consider Column 1 to be my preferred specification, as it is the most parsimonious.

In addition to my "small group" outcome, I also examine the effects of differential reimbursement across the entire distribution of selected group sizes. Panel A of Figure 6 displays the raw cumulative distributions of selected practice sizes separately for the high- and low-reimbursement sides of the border. (Note that the x-axis utilizes a logarithmic scale of base 2, which is employed because the group sizes exhibit substantial positive skewness.) Meaningfully, the distribution on the high-payment side of the border stochastically dominates the low-payment distribution, indicating that a much higher fraction of primary care physicians in higher-payment areas choose smaller practices. A Kolmogorov-Smirnov test easily rejects that the two sides have equal distributions ($p < 0.01$).

To evaluate this within the context of my regression discontinuity framework, I estimate sep-

[22]See Appendix Figure A4 for illustration of these physician clusters and Appendix Section B for a discussion of how they were constructed using k-means clustering. To further assess the role of spatial characteristics, Appendix Figure A3 presents estimates and associated regression discontinuity plots of a specification *without* any spatial fixed effects and finds results that are qualitatively similar, though slightly larger in magnitude.

arate RD-IV regressions (Equations 10 and 11) for various points in the cumulative distribution. This is done by utilizing an outcome of $Y_{ip} = 1(GroupSize \leq g)$, where g represents a point in the cumulative distribution.[23] The estimates from this analysis, along with their associated 90% confidence intervals are presented in Figure 6, Panel B.[24] This figure supports the conclusions drawn from Panel A: in response to higher payment, providers are more likely to sort into the small group practices and less likely to be employed by the largest groups.

The effects on practice size manifest very early in the distribution of group sizes. The RD-IV estimate implies that, in response to reimbursement that is 1 percent more generous, there is a 1.8 p.p. increase ($p = 0.050$) in primary care physicians who are affiliated with practice groups with 1 or 2 providers. The effects on the smallest group persist throughout the distribution until they are offset by reductions in groups of 150 greater ($\delta_{RD-IV} = -1.9$, $p = 0.075$).[25] Groups of this size are generally comprised of hospital systems or physician groups that are affiliated with a particular hospital system. Accordingly, to frame my results another way, pricing *pressure* might force providers who would have non-financial preferences to practice in a small group to select larger organizations, with the net effect of increasing primary care physicians employed by large, vertically integrated health systems and their associated physician employment groups. More generally, these results may suggest that diminished profitability is a driving force behind the recent trend of consolidation of small physician practices into larger ones (e.g., Capps et al., 2017; Kane, 2019; Muhlestein and Smith, 2016).

Finally, I consider two other measures of practice structure, both of which are presented in Appendix Figure A8. The first measure, presented in Panel A, details effects of differential effects on practice *location* size. Specifically, the outcome for this analysis is the log number of providers at a given practice location, defined as an address-group combination. I do not detect any effects on practice size, suggesting that providers are not responding by *spatially* consolidating into locations with more physicians. This is consistent with the lack of spatial response detailed at length in Section 4.2. The second measure, presented in Panel B, details the impacts of differential reimbursement on average log group size. This measure is not presented as a primary outcome because it is challenging to interpret and because moderate shifts in the *distribution* of group sizes can have very large effects on *average* group size. The results in Panel B underscore this: the discontinuity displayed implies a decrease of 0.54 log points (roughly a 41% decrease) in group size. While this may appear implausible at first, approximately the same result would be achieved if 5% of the primary care providers in a market moved from the 75th percentile of group size (roughly 250 providers) to the 25th percentile of group size (roughly 2 providers). Based on the differences in cumulative distributions displayed in Figure 6, effects of roughly this magnitude appear to be present, supporting the assertion that the effects on average size are reasonable.[26]

[23] This analysis is performed using similar methodology as Barcellos et al. (2019), who examine the distributional effects of education on health within a regression discontinuity framework.

[24] For an analogous exercise focusing on the probability distribution, rather than cumulative distribution, see Appendix Section C and Appendix Figure A5.

[25] Regression discontinuity plots for these figures are displayed in Appendix Figure A6.

[26] I also consider two other more frequently examined outcomes from the physician payment literature: per-patient

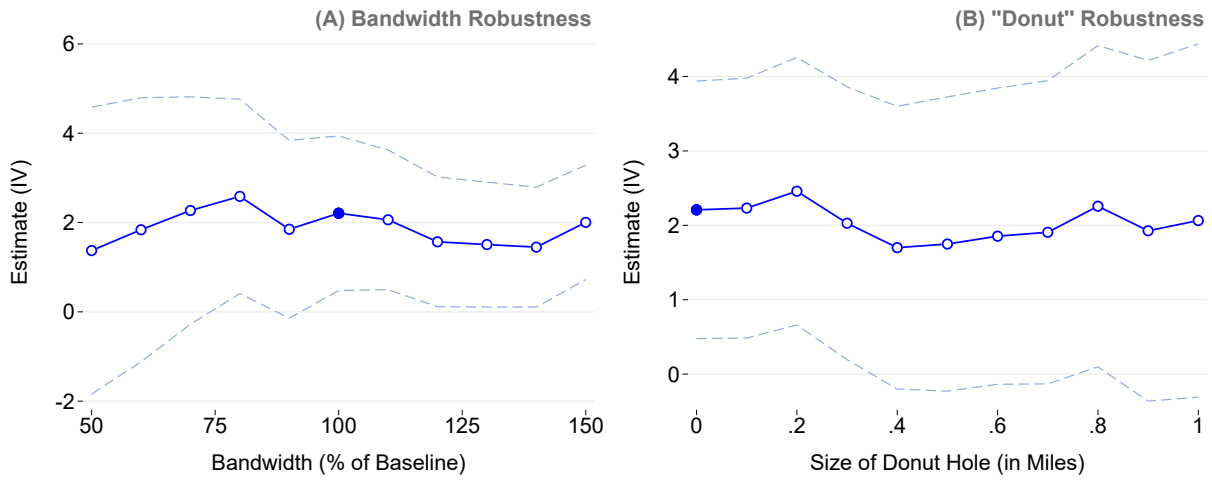


Figure 7: Robustness: Bandwidth Sensitivity

Notes: The purpose of this figure is to display robustness tests of the main result displayed in Table 1. Panel A displays point estimates for the RD-IV specification for separate regressions utilizing varying bandwidths (ranging from 50% to 150% of the baseline bandwidth, or 2 to 6 miles). Panel B shows estimates from the RD-IV specification for separate "donut" regressions on samples that omit observations within a given distance from the cutoff, ranging from 0 to 1 mile(s). In both panels, 90% confidence intervals are displayed using dashed lines and the preferred estimate is shaded dark blue. These analyses utilize the High-Impact Sample discussed in Section 2.2.

Source: Author calculations using CMS Physician Fee Schedule, CMS Physician Compare Data, CMS Physician and Other Supplier Public Use File.

5.2 Robustness

Aside from the assumptions that unobservable characteristics do not vary over the boundary and that there is no meaningful sorting by providers, the choice of bandwidth is the most critical to my analysis. To consider sensitivity of my results to this assumption, I re-estimate my RD-IV results (Equations 10 and 11) for various multiples of bandwidth, as shown in Figure 7 Panel A. Overall, the coefficients are highly consistent, although they lose precision at narrower bandwidths, which is consistent with the bias-variance trade-off.[27]

Additionally, motivated by the outlier point immediately to the right of the cutoff in Figure 5, I perform a "donut" hole analysis (Figure 7, Panel B), by estimating separate regressions on samples that omit observations within a given distance from the cutoff, ranging from 0 to 1 mile(s). As displayed in the figure, the RD-IV estimates are highly consistent over the size of the omitted sample, although my estimates do predictably fail to attain traditional levels of statistical significance when increasingly large number of observations are omitted. In addition to measuring sensitivity to outliers, this test accomplishes two goals. First, insofar as there is sorting *immediately* around

resource utilization (i.e. the intensive-margin treatment decision) and the number of Medicare patients treated (i.e. the extensive-margin treatment decision). See Appendix Section C and Appendix Figure A7 for further discussion.

[27] An alternate approach is to utilize methods developed by Calonico et al. (2014) to non-parametrically select bandwidth and estimate results within a local linear regression specification. When applying these methods within a fuzzy regression discontinuity framework (comparable to my RD-IV specification), I obtain highly similar results to my baseline analysis ($\delta = 2.205$, $p = 0.016$).

the cutoff that is not otherwise detected by my tests in Section 4.2, physicians who have manipulated the running variable (and therefore unobservably differ across the cutoff) will be omitted from my sample. Second, these analyses will reduce measurement error in my treatment variable from inaccurately geocoded physician locations, to the degree that such measurement error exists. The consistency of the estimates over the range of "donut hole" sizes supports the idea that my estimates are not meaningfully biased by endogenous sorting or measurement error.

I perform a variety of other robustness checks, which are detailed and discussed further in Appendix Section C. However, I will briefly summarize them here. First, I consider whether results are being disproportionately influenced by individual locality borders by re-running my preferred specification while omitting a given border from the analysis. The results of this analysis, which are detailed in Appendix Figure A9, demonstrate that my results are highly consistent when omitting any given border from my sample. Next, I consider sensitivity of my estimates to different choices of clustering strategy, kernel, and controls (i.e., using quadratic and border-varying linear distance controls as well as granular physician specialty controls). I also consider a wide variety of sample restrictions to test sensitivity of my estimates to sample construction choices and spatial idiosyncrasies. The results of these tests are summarized in Appendix Figure A10, where I find that my point estimates are consistent regardless of specification choice.

5.3 Alternate IV and Full Sample Results

The results so far are based the on reduced-form and RD-IV analyses discussed in Section 4. While these methods have the advantage of transparency (i.e. both the first and second stages can be clearly displayed graphically), they do not utilize all the plausibly exogenous variation that is generated by these border discontinuities. Specifically, the RD-IV design utilizes the *average* discontinuity across all borders, rather than exploiting discontinuities specific to each border. To address this issue, I estimate an alternative design, where the second stage is unchanged from Equation 10, but the first-stage equation utilizes multiple discontinuities and allows the θ coefficient to vary across borders:

$$Y_{ip} = \delta \cdot \widehat{GAF}_{ip} + \phi_{b(p)} + f(Z_p, b(p)) + \varepsilon_{ip}, \tag{12}$$

$$GAF_{ip} = \sum_{k \in \mathcal{B}} \theta_k \cdot High_p \times 1[b(p) = k] + \phi_{b(p)} + f(Z_p, b(p)) + \varepsilon_{ip}. \tag{13}$$

The instruments in Equation 13 are created by interacting (i) a binary variable equal to 1 if a provider is located on the high-payment side of the border and (ii) an indicator variable for assignment to a given border. I also allow trends in the running variable to vary on each side of the cutoff *and* by each border, so the first-stage equation is effectively estimating a separate regression discontinuity for each of the borders in my sample. I refer to this estimation strategy as the "RD-Border" specification.

I apply this strategy, along with my RD-IV specification, to both my High Impact Sample (i.e.,

Table 2: Full-Sample and Alternate IV Results

Outcome	(1)		(2)		(3)		(4)	
% of Physicians in Small Groups	2.208	**	1.339		1.703	**	1.673	**
	(1.052)		(1.709)		(0.849)		(0.850)	
Sample	High Impact		Full		High Impact		Full	
Instrumental variable	RD-IV		RD-IV		RD-Border		RD-Border	
Observations	5,065		9,685		5,065		9,685	
First stage F-statistic	399		148		7,468		5,920	

Notes: This table reports results for different samples and instrumental variable strategies. Columns 1 and 2 display coefficients and associated standard errors when estimating using the RD-IV (Equations 10 and 11), whereas Columns 3 and 4 display results for the RD-Border IV (Equations 12 and 13). Columns 1 and 3 utilize the High-Impact Sample discussed in Section 2.2 while Columns 2 and 4 use all available borders. All specifications utilize a rectangular kernel in distance from the boundary and are clustered at the Census-tract level. Significance at 10%, 5%, and 1% levels is represented by *, **, and ***, respectively.

Source: Author calculations using CMS Physician Fee Schedule, CMS Physician Compare Data, CMS Physician and Other Supplier Public Use File.

those borders with substantial payment differences) and the full sample of borders. The results are displayed in Table 2. The first column repeats the estimates from my preferred specification (Table 1) for comparison. When the RD-IV estimation is applied to the full sample—as shown in Column 2—the estimate decreases meaningfully and also suffers from a high degree of imprecision. This is consistent with a first stage that is roughly half as large as when the High-Impact Sample is used (Appendix Figure A11). Columns 3 and 4 repeat the analyses of Columns 1 and 2 when applying the "RD-Border" estimation strategy to potentially increase the precision of my estimates. When this method is applied to the High-Impact Sample (Column 3), estimates are similar to my RD-IV analysis (although somewhat smaller, they are not statistically different). When I apply the RD-Border estimation to the full sample, the results—displayed in Column 4—yield nearly identical estimates to those from the High-Impact sample. Accordingly, these analyses demonstrate that the results I find are not driven by my choice of sample selection.[28]

5.4 Heterogeneity

I conclude this section by conducting two heterogeneity analyses to better understand the effect of Medicare pricing on physician practice choice. First, because the financial impact of Medicare reimbursement policy will likely be larger for physicians practicing in areas with a higher share of Medicare-eligible population (those aged 65 and over), I test whether PCPs practicing in high-share areas respond more strongly to differential Medicare reimbursement. To do so, I divide my High-Impact Sample into two groups: primary care physicians assigned to borders with above-median shares of Medicare-eligible population and PCPs assigned to borders with below-median

[28] I also re-perform my distributional analyses from the previous section using the full sample and RD-Border estimation. These results, which are displayed in Appendix Figure A12, are consistent with those from my High-Impact Sample analysis.

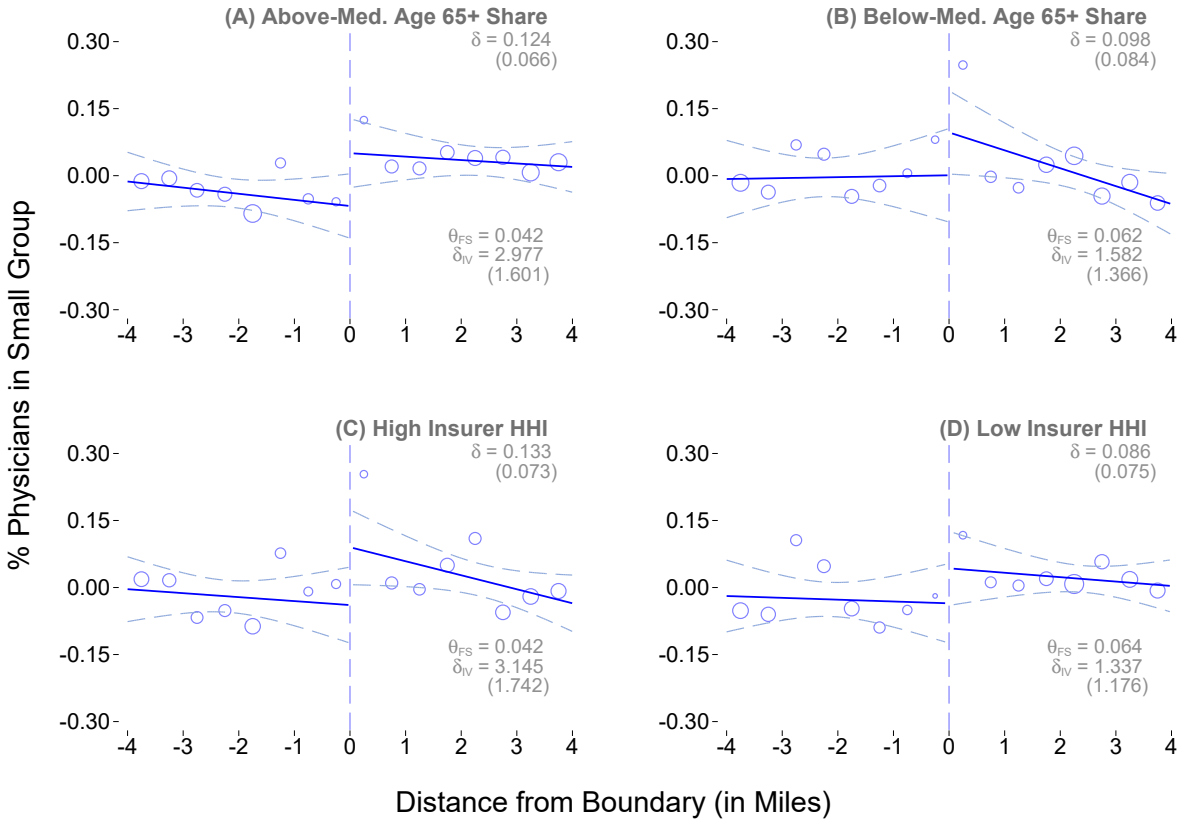


Figure 8: Heterogeneous Effects by Border Characteristics

Notes: The purpose of this figure is display heterogeneity of results by physicians' local environment. Within each panel, reduced-form estimates (Equation 8) and associated standard errors are in the upper-right corner, while the first-stage (Equation 11) and RD-IV estimate (δ_{IV}, from Equation 10) and associated standard errors are in the bottom-right corner. The first set of estimates, displayed in Panels A and B, splits the sample according to borders that have above- and below-median shares of over Medicare-eligible adults, respectively. The second set of estimates, displayed in Panels C and D, splits the sample according to borders that have above- and below-median levels of health insurer market concentration, as measured by the Herfindahl-Hirschman Index ("HHI").

Source: Author calculations using CMS Physician Fee Schedule, CMS Physician Compare Data, CMS Physician and Other Supplier Public Use File, Census Block Group 5-year ACS Estimates (Ruggles et al., 2019), American Medical Association Market Share Calculations.

shares.[29] The results of this analysis are displayed in Figure 8, Panels A and B, which mirror regression discontinuity plots presented in this paper, but with additional information about the first stage (θ_{FS}) and RD-IV estimates (δ_{IV}) presented in the bottom-right corner. As expected, estimates are substantially larger for primary care doctors who practice in areas where there are more Medicare-eligible individuals, and who are therefore more dependent on Medicare reimbursement. While it is possible that these effects are driven by other factors that are correlated with both larger over-age-65 population shares and higher degrees of responsiveness to payment differentials, this analysis provides further suggestive evidence that the differences in practice choice

[29] I choose to divide providers into groups based on border-level, rather than individual-level, characteristics because it is possible that the number of Medicare patients treated is endogenous.

are driven by Medicare's reimbursement decisions.[30]

Next, I investigate how variation in private insurer bargaining power (as proxied by market concentration) may affect physician responses to differential Medicare reimbursement. The effects of insurer bargaining power are *ex-ante* ambiguous. In the context of my model (Equation 4), powerful insurers are more likely to push negotiated private prices, $p^o(s)$, closer to the physician's reserve price, p^m (in other words, $\theta(s) \rightarrow 1$). This would imply that physicians in more highly concentrated private insurance markets would be more responsive to differential Medicare reimbursement, as Medicare differentials translate more directly to private prices. However, as noted in Barrette et al. (2020), countervailing market power by insurance companies may offset the increases in price that come from higher concentration in the provider markets. This means that the financial returns from selecting a larger group may be lower in markets with powerful insurance companies, and thus it is possible that physician practice choice may be *less* responsive in High-HHI insurance markets.[31] To test which effect dominates, I divide my High-Impact Sample into physicians who are assigned to borders with above- and below-median market concentration and perform separate regressions on each sample.[32] The results, displayed in Figure 8 Panels C and D, show that physicians in highly concentrated markets are more responsive to Medicare pricing policy, suggesting that the tighter relationship of private and Medicare prices in these markets may dominate any loss in incremental bargaining power gains from consolidation.

6 Discussion

This paper finds that Medicare payments affect practice structure choice among primary care physicians. Specifically, I find that a 1 percent higher payment rate leads to a 1.7-2.2 p.p. increase in the fraction of primary care doctors affiliated with small practice groups. Further investigation of practice size distribution finds that a 1 percent higher reimbursement rate increases the share of doctors in 1-2 provider groups by 1.8 percentage points and *decreases* the share in very large groups ($\geq$ 150 providers) by 1.9 percentage points.

While addressing the welfare impact of these changes is beyond the scope of this paper, it is worthwhile to discuss the implications of larger group sizes on private prices for medical services,

[30] In a related analysis, I also separately consider responses for the two largest primary care specialties: Internal Medicine and Family Practice. Given that Internal Medicine physicians generally have a larger share of Medicare beneficiaries (roughly 49% more per the 2010-2015 National Ambulatory Medical Care Surveys), it is expected that they would exhibit a stronger response to differential reimbursement. The results of the analyses displayed in Appendix Figure A13 Panels A and B, confirm this: the RD-IV point estimate is approximately 61% larger for Internal Medicine physicians than for Family Practice doctors (although it should be noted that these estimates are not statistically different at traditional levels). I also consider other types of physician heterogeneity (Panels C through H of Figure A13) and find little other evidence of differential response. See Appendix Section C for further discussion.

[31] This would be formally characterized by $\theta_s^{High\ HHI} \geq \theta_s^{Low\ HHI}$, where values of θ_s that are *more* negative represent larger shifts toward the insurers' reservation price.

[32] Market concentration is determined using CBSA-level measures of HHI obtained from American Medical Association (2016). These CBSA-level estimates are mapped to the border level via a weighted average based on the number of physicians within a bandwidth-distance of the border that are in a given CBSA.

achieved through higher levels of bargaining power.[33] To do so, I relate select measures of group size to the Herfindahl-Hirschman Index ("HHI"), the most common metric used to study market power.[34] I begin by constructing a measure of HHI by using a similar method to Dunn and Shapiro (2014), who use fixed travel-time HHI ("FTHHI") to assess market power in physician markets. I calculate the market share for each Census tract by first weighting each provider based on a 20-minute travel-time distance from the population-weighted centroid of that Census tract.[35] I then aggregate providers within the same group, and divide by the total number to obtain each group's Census-tract level market share:

$$s_g = \frac{\sum_{j \in J(g)} w_j}{\sum_{f \in \mathcal{F}} \sum_{j \in J(f)} w_j}, \tag{14}$$

where j indicates a provider and w_j the associated weight; $J(g)$ represents the set of providers affiliated with group g; and $\mathcal{F}$ is the set of all firms in the market. These shares are then used to calculate the FTHHI for each Census tract, which form the basis of the population-weighted average FTHHI for each locality border.

I then relate the mean FTHHI to two measures: (i) the market-level fraction of providers in small groups (the main outcome); and (ii) the market-level fraction of providers in groups of 1-2. The latter measure was chosen because it is the best predictor of FTHHI among my statistically significant distributional results (see Appendix Figure A14). There is a strong linear relationship between group size measures and FTHHI: a 1 p.p. increase in the fraction of small groups corresponds to a 2.2% *decrease* in FTHHI, whereas a 1 p.p. increase in the fraction of 1-2 provider groups corresponds to a 5.4% decrease in FTHHI. Using these relationships, along with the estimated price increase from a 1% increase in FTHHI from Dunn and Shapiro (2014) and the estimates from Section 5, I calculate that a 1% *decrease* in Medicare prices induces a 0.15% - 0.50% increase in private market prices due to increased market power.[36] Given the size of the privately financed portion of the primary care services market ($60 billion), this translates to a $90 - $290 million annual transfer from insurance companies to medical providers, compared to $280 million in cost savings for Medicare, before accounting for other margins of response to Medicare payment.[37] Additionally,

[33] Given the changes in physician responses, welfare analysis would require an evaluation of the health impacts, if any, stemming from those responses. As noted in the introduction, the evidence on the health benefits of larger and vertically integrated practices is mixed, with some (Epstein et al., 2010; Baker et al., 2019) arguing that larger practices have health benefits, while others do not find any evidence of health improvements (e.g., Scott et al., 2017) or negative health effects (Short and Ho, 2019).

[34] I am unable to directly assess the impact of differential Medicare payments on HHI because it is a market-level measure, whereas my treatment and control providers are all in the same market by construction.

[35] Weighting for provider j is done according to the formula, $w_j = \max\{1 - (d_j/\bar{d}), 0\}$, where d_j is the distance of the provider from the centroid and $\bar{d}$ is the 20-minute travel-time distance. The equivalent distance for a trip of 20 minutes was calculated using data from the NHTS.

[36] Dunn and Shapiro (2014) estimate that a 1% increase in 20-minute FTHHI results in a 0.04% to 0.05% increase in private prices.

[37] The size of the private and Medicare markets were derived by multiplying total spending on personal health care (CMS, 2018) by the estimated fraction of total spending attributable from primary care (Jabbarpour et al., 2019). In addition to holding other responses fixed, this calculation also makes the critical assumption that my results are generalizable to a nationwide decrease in Medicare payments. It is also possible that the transfer is overstated, as my calculations do not account for lower costs from larger groups that may be passed onto insurers.

Baker et al. (2014) demonstrate that increased vertical integration is associated with higher hospital prices when hospitals own physician practices. Accordingly, to the extent that my estimated effects on physicians employed by the largest groups represents a change in hospital ownership of physician practices, price decreases could also have substantial cost spillovers in the inpatient market as well.[38]

In summary, changes in practice structure stemming from lower Medicare prices have the potential to lead to higher private-market prices for both physician and hospital services. Depending on the incidence of these price changes, a substantial degree of savings from cuts in Medicare reimbursement could be borne by privately insured individuals in the form of higher out-of-pocket medical costs and health insurance premiums.

[38]Given the size of the private hospital market ($427B per CMS, 2018), even small effects on vertical integration could impose large spillovers in relation to cost savings by Medicare. For instance, given that a 1 p.p. increase in vertical integration results in 0.137% increase in price (Appendix Table 2 of Baker et al., 2014), only a 0.47 p.p. increase in vertically integrated physicians would be required to match the naive Medicare cost savings from a 1% decrease in Medicare prices. Such an estimate is well within the range of possibility, based on the results displayed in Figure 6.

References

Acemoglu, D. and A. Finkelstein (2008). Input and Technology Choices in Regulated Industries: Evidence from the Health Care Sector. *Journal of Political Economy 116*(5), 837–880.

Alexander, D. and M. Schnell (2019). The Impacts of Physician Payments on Patient Access, Use, and Health. Working Paper 26095, National Bureau of Economic Research.

Ambrus, A., E. Field, and R. Gonzalez (2020, February). "Loss in the Time of Cholera: Long-run Impact of a Disease Epidemic on the Urban Landscape". *American Economic Review 110*(2), 475–525.

American Medical Association (2016). Competition in Health Insurance: A Comprehensive Study of U.S. Markets, 15th Edition.

Baker, L. C., M. K. Bundorf, and D. P. Kessler (2014). Vertical Integration: Hospital Ownership Of Physician Practices Is Associated With Higher Prices And Spending. *Health Affairs* (5), 756–763.

Baker, L. C., M. K. Bundorf, and A. B. Royalty (2019). The Effects of Multispecialty Group Practice on Health Care Spending and Use. Working Paper 25915, National Bureau of Economic Research.

Barcellos, S. H., L. S. Carvalho, and P. Turley (2019). Distributional Effects of Education on Health. Working Paper 25898, National Bureau of Economic Research.

Barrette, E., G. Gowrisankaran, and R. Town (2020). Countervailing Market Power and Hospital Competition. Working Paper 24354, National Bureau of Economic Research.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica 82*(6), 2295–2326.

Capps, C., D. Dranove, and C. Ody (2017). Physician Practice Consolidation Driven By Small Acquisitions, So Antitrust Agencies Have Few Tools To Intervene. *Health Affairs 36*(9), 1556–1563.

Casalino, L., K. Devers, T. Lake, M. Reed, and J. Stoddard (2003). Benefits of and Barriers to Large Medical Group Practice in the United States. *Archives of Internal Medicine 163*(16), 1958–1964.

Center for Medicare and Medicaid Services (2018). Historical National Health Expenditure Accounts. `https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/Downloads/Tables.zip`. Accessed 5/3/2018.

Centers for Medicare and Medicaid Services (2015). Physician Value-Based Payment Modifier and the Physician Feedback Program. `https://www.cms.gov/newsroom/fact-sheets/physician-value-based-payment-modifier-and-physician-feedback-program`. Accessed 8/2/2020.

Clemens, J. and J. D. Gottlieb (2014). Do Physicians' Financial Incentives Affect Medical Treatment and Patient Health? *The American Economic Review 104*(4), 1320–1349.

Clemens, J. and J. D. Gottlieb (2017). In the Shadow of a Giant: Medicare's Influence on Private Physician Payments. *Journal of Political Economy 125*(1), 1–39.

Craig, S., M. Grennan, and A. Swanson (2018). Mergers and Marginal Costs: New Evidence on Hospital Buyer Power. Working Paper 24926, National Bureau of Economic Research.

Dell, M. (2010). The Persistent Effects of Peru's Mining Mita. *Econometrica 78*(6), 1863–1903.

Dranove, D. and C. Ody (2019). Employed for Higher Pay? How Medicare Payment Rules Affect Hospital Employment of Physicians. *American Economic Journal: Economic Policy 11*(4), 249–71.

Dunn, A. and A. Shapiro (2014). Do Physicians Possess Market Power? *The Journal of Law and Economics 57*(1), 159–193.

Epstein, A. J., J. D. Ketcham, and S. Nicholson (2010). Specialization and Matching in Professional Services Firms. *The RAND Journal of Economics 41*(4), 811–834.

Finkelstein, A. (2007). The Aggregate Effects of Health Insurance: Evidence from the Introduction of Medicare. *The Quarterly Journal of Economics 122*(1), 1–37.

Glied, S. and J. G. Zivin (2002). How Do Doctors Behave When Some (But Not All) of Their Patients Are in Managed Care? *Journal of Health Economics 21*(2), 337 – 353.

Gowrisankaran, G., A. Nevo, and R. Town (2015). Mergers When Prices Are Negotiated: Evidence from the Hospital Industry. *American Economic Review 105*(1), 172–203.

Jabbarpour, Y., A. Greiner, A. Jetty, M. Coffman, C. Jose, S. Petterson, K. Pivaral, R. Phillips, A. Bazemore, and A. N. Kane (2019). Investing in Primary Care: A State-Level Analysis. Technical report, Patient-Centered Primary Care Collaborative and the Robert Graham Center.

Kane, C. K. (2019). Updated Data on Physician Practice Arrangements: For the First Time, Fewer Physicians are Owners Than Employees. Technical report, American Medical Association.

Kessler, D. P. and M. B. McClellan (2000). Is Hospital Competition Socially Wasteful? *The Quarterly Journal of Economics 115*(2), 577–615.

Lipton, W. (2012). Is a Bigger Medical Practice Always Better? `http://www.physicianspractice.com/overhead/bigger-medical-practice-always-better`. Accessed 5/3/2018.

McCrary, J. (2008). Manipulation of the Running variable in the Regression Discontinuity Design: A Density Test. *Journal of Econometrics 142*(2), 698–714.

McGuire, T. G. and M. V. Pauly (1991). Physician Response to Fee Changes with Multiple Payers. *Journal of Health Economics 10*(4), 385 – 410.

Muhlestein, D. B. and N. J. Smith (2016). Physician Consolidation: Rapid Movement From Small To Large Group Practices, 2013–15. *Health Affairs 35*(9), 1638–1642.

Nicholson, S. and C. Propper (2012). Medical Workforce. In M. V. Pauly, T. G. McGuire, and P. P. Barros (Eds.), *Handbook of Health Economics*, Volume two, Chapter 14, pp. 873–925. Amsterdam: North Holland.

Ruggles, S., S. Flood, R. Goeken, J. Grover, E. Meyer, J. Pacas, and M. Sobek (2019). IPUMS USA: Version 9.0 [dataset]. Minneapolis, MN. https://doi.org/10.18128/D010.V9.0.

Schnell, M. and J. Currie (2018). Addressing the Opioid Epidemic: Is There a Role for Physician Education? *American Journal of Health Economics 4*(3), 383–410.

Scott, K. W., E. J. Orav, D. M. Cutler, and A. K. Jha (2017). Changes in Hospital–Physician Affiliations in U.S. Hospitals and Their Effect on Quality of Care. *Annals of Internal Medicine 166*(1), 1–8.

Short, M. N. and V. Ho (2019). Weighing the Effects of Vertical Integration Versus Market Concentration on Hospital Quality. *Medical Care Research and Review*, 1–18. PMID: 30741109.

Xierali, I. M. (2018). Physician Multisite Practicing: Impact on Access to Care. *Journal of the American Board of Family Medicine 31*(2), 260–269.

Zuckerman, S., L. Skopec, and M. Epstein (2017). Medicaid Physician Fees after the ACA Primary Care Fee Bump. Technical report, Urban Institute.

Appendix: for Online Publication

A Variations of the Theoretical Model

The model in Section 3 makes several simplifications for purposes of clarity and tractability. Accordingly, the purpose of this Appendix Section is to test the model's sensitivity to these assumptions, particularly those regarding intensive-margin quantity responses to differential prices. This will be done in two ways: (1) by maintaining the original model, with an additional assumption that Medicare quantity responds positively to price differences and (2) deriving a model with joint choices of practice size and Medicare quantity.

Preliminaries. Recall the following equations from Section 3:

$$\max_s u(s) = v(\pi(s)) - w(s) \text{ subject to} \tag{A1}$$

$$\pi(s) = p^o(s) \cdot q^o + p^m \cdot q^m - mc(s) \cdot (q^m + q^o) \tag{A2}$$

$$p^o(s) = (1 - \theta(s)) \cdot p^{ins} + \theta(s) \cdot p^m, \tag{A3}$$

which produce the comparative static of Medicare prices, p^m, on group size, s:

$$\frac{ds}{dp^m} = -\frac{v_{\pi\pi}\pi_s\pi_{p^m} + v_\pi \pi_{sp^m}}{u_{ss}} < 0 \tag{A4}$$

The sign of the equation above is based the increasing and concave nature of utility, as well as the following conditions derived from the model:

$$\begin{aligned} \pi_s &= p^o_s q^o - mc_s(q^m + q^o) \\ &= -\theta_s(p^{ins} - p^m)q^o - mc_s(q^m + q^o) > 0 \end{aligned} \tag{A5}$$

$$\pi_{p^m} = \theta(s)q^o + q^m > 0 \tag{A6}$$

$$\pi_{sp^m} = \theta_s q^o < 0. \tag{A7}$$

Original Model + Increasing Quantity Assumption. To consider how quantity responses might affect model predictions, first I allow Medicare quantities to respond positively to Medicare prices (formally, $q^m_p > 0$), as shown empirically in Clemens and Gottlieb (2014).[1] Such an assumption would leave $v_{\pi\pi}, v_\pi, u_{ss}$, and π_s unchanged. The affected terms would now include:

$$\pi_{p^m} = \pi^{orig}_{p^m} + (p^m - mc)q^m_p > 0 \tag{A8}$$

$$\pi_{sp^m} = \pi^{orig}_{sp^m} - mc_s q^m_p \gtrless 0, \tag{A9}$$

where $\pi^{orig}_{p^m}$ and $\pi^{orig}_{sp^m}$ represent the values in Equations A6 and A7, respectively. The sign for π_{p^m} is the same as in the original model, under the weak assumption that the marginal Medicare patient is profitable ($p^m - mc > 0$). However, the sign for π_{sp^m} is now technically ambiguous, as the term

[1] Private quantities are assumed to be unresponsive to Medicare prices for clarity of exposition.

$-mc_s q_p^m$ is positive. Despite the ambiguous term, it is unlikely that this term is sufficiently influential as to change model predictions due to its small probable magnitude. Empirical evidence on cost savings from mergers and organizational size increases suggest (at best) modest cost savings (see, for instance, Craig et al. 2018). When these marginal cost savings are interacted with known intensive-margin quantity responses (with elasticities of approximately 1-2%), they are unlikely to result in decision-altering quantities. Accordingly, the predictions of the augmented model are likely unchanged from the baseline.

Model with Endogenous Choice of Quantity. I next expand the model to allow physicians to choose the intensive-margin quantity of services delivered to patients (which implicitly assumes that physicians induce demand as agents of patients). In response to changing quantity, physicians also incur non-financial disutility, represented by $h(q)$, due to the cost of inducement. Formally, the physician solves:

$$\max_{s,q^m} u(s,q^m) = v(\pi(s,q^m)) - w(s) - h(q^m) \text{ subject to} \tag{A10}$$

$$\pi(s,q^m) = p^o(s)\cdot q^o + p^m \cdot q^m - mc(s)\cdot(q^m + q^o) \tag{A11}$$

$$p^o(s) = (1-\theta(s))\cdot p^{ins} + \theta(s)\cdot p^m, \tag{A12}$$

where the assumptions from the previous model remain, except that $w(s)$ is assumed to be linear (for simplicity) and $h(q)$ is assumed to be increasing in q.[2] The first order conditions of this model are:

$$\frac{\partial u(s)}{\partial s} = 0 \Leftrightarrow v_\pi \pi_s - w_s = 0 \tag{A13}$$

$$\frac{\partial u(s)}{\partial q^m} = 0 \Leftrightarrow v_\pi \pi_{q^m} - h_{q^m} = 0. \tag{A14}$$

Combining first order conditions to obtain:

$$\phi(s,q^m) = w_s \pi_{q^m} - h_{q^m}\pi_s \tag{A15}$$

and totally differentiating to obtain the comparative static of Medicare prices, p^m, on group size, s, yields:

$$\frac{ds}{dp^m} = -\frac{\partial\phi(s,q^m)/\partial p^m}{\partial\phi(s,q^m)/\partial s} = -\frac{w_s\pi_{q^m p^m} - h_{q^m}\pi_{sp^m}}{w_{ss}\pi_{q^m} + w_s\pi_{sq^m} - h_{q^m}\pi_{ss}}. \tag{A16}$$

Based on the assumptions of the model, signs of each component of the comparative static are as

[2] As before, private quantities are fixed for clarity of exposition.

follows:

$$w_s, h_{q^m} > 0 \tag{A17}$$

$$w_{ss} = 0 \tag{A18}$$

$$\pi_{q^m p^m} = 1 \tag{A19}$$

$$\pi_{sp^m} = \theta_s q^o < 0 \tag{A20}$$

$$\pi_{sq^m} = -MC_s > 0 \tag{A21}$$

$$\pi_{ss} = p_s^o q^o - mc_{ss}(q^m + q^o) \tag{A22}$$

$$= -\theta_{ss}(p^{ins} - p^m)q^o - mc_{ss}(q^m + q^o) > 0. \tag{A23}$$

Accordingly, the sign of Equation A16 is:

$$\frac{ds}{dp^m} < 0, \tag{A24}$$

which is the same prediction given by the original model.

B Sample Construction

The primary datasets used for my analysis are as follows:

1. **Medicare Physician and Other Supplier Public Use File**. This dataset supplies provider-level data for all providers who treat more than 10 Medicare patients in a given year. The data includes information detailing the exact addresses, summary information on patients treated—including average age and average hierarchical condition category score, which measures patient health—and provider characteristics such as certification (ex. Medical Doctor, Doctor of Osteopathic Medicine) and primary specialty type. The detail file also includes service-level detail regarding procedures by the provider in a given year. Relative value units were assigned to each service by merging in the 2015 Physician Fee Schedule's Relative Value files obtained from CMS. The source data for utilization and patient-population characteristics are derived from the CMS Chronic Conditions Warehouse, a database with 100% of Medicare enrollment and fee-for-service claims data. Provider characteristics and address are obtained from the National Plan and Provider Enumeration System ("NPPES"). According to NPPES guidelines, any address changes must be communicated by providers within 30 days of the change.

2. **Medicare Physician Compare Data**. This dataset is used to gather physician-level information regarding group practice affiliation, the size of associated medical groups, medical school attended, and year of medical school graduation. Medical school rankings were obtained using the methodology detailed in Schnell and Currie (2018), and provider experience variables were generated by subtracting the current year (2015) from the date of medical school graduation. The source data for Physician Compare variables is the Provider Enrollment, Chain, and Ownership System ("PECOS"), which is Medicare's enrollment and revalidation system. Group practice affiliation is determined based on the tax identification number that is associated with a given physician's NPI. Accordingly, it is possible for providers to have multiple affiliations in the Physician Compare data, which is true for roughly 14% of providers in my sample. I assign providers to their smallest practice under the assumption that larger practices are more likely to be included only due to practice privileges and are therefore not the providers' main practice. As discussed in Section C, this assumption does not meaningfully affect my estimates.

3. **Medicare Physician Fee Schedule**. This dataset provides detail regarding geographic adjustment factors by locality, RVUs-by-HCPCS code, and zip-to-locality crosswalks. I use the zip-to-locality crosswalks to form locality borders, defined by adjacent zip codes that are assigned different geographic adjustment factors. Physicians are paid using the geographic adjustment factors that correspond to their practice address.

4. **Census Data**. I obtained local data on the Census-block-group level using 2011-2015 ACS 5-year estimates from the U.S. Census Bureau and IPUMS (Ruggles et al., 2019). This data

includes the following variables, which were used as "local" controls in my analysis:

(a) Percentage of population in given age bins (less than 65, 65-69, 70-74, 75-79, 80-84, and 84-plus);

(b) Percentage of population by education level (less than high school, high school, some college, college or more);

(c) Percentage of population by insurance coverage (uninsured, Medicare only, Medicaid only, other insurance only, Medicare+Medicaid, Medicare+Private Insurance);

(d) Median household income;

(e) Median home value; and

(f) Mean population density.

5. **Spatial Data**. Geographic shapefiles (State- and ZCTA-level) were obtained from the U.S. Census.

Using this data, the main sample construction was as follows:

1. The ZCTA-level shapefiles and zip-to-locality crosswalk were utilized to create locality geographies in ArcGIS. Locality borders were then generated for each combination of within-state adjacent localities (i.e. no cross-state borders were used). Borders that corresponded with large rivers were also omitted, as other characteristics would likely not evolve smoothly across the boundary in these instances.

2. Physician addresses from the Physician and Other Supplier Public Use File were geocoded using Texas A&M GeoServices. Then the shortest straight-line distance was calculated between each border and candidate border point within 10 miles.

3. The Physician and Other Supplier Public Use Variables and Physician Compare variables were combined. Only primary care physicians were retained.[3] Next, I dropped all physicians that had outlier geocoordinates (i.e. those outside of the boundaries of their listed zip code) and physicians who practice in nine-digit zip codes that had a different locality than their respective five-digit zip code were also dropped. Finally, I keep only providers who are within one bandwidth of the border.

4. Only borders that have 5 providers on each side of the boundary were maintained.

5. I also created "clusters" of physicians in each border to more finely control for the spatial characteristics of providers (see Table 1, Column 4). To do so, I utilized k-means clustering *separately for each border* based on physician longitude and latitude. The algorithm was set

[3]Primary care physicians included those who listed their specialties as: Family Practice, General Practice, Geriatric Medicine, Internal Medicine, Osteopathic Manipulative Medicine, Pediatric Medicine, or Preventive Medicine. Among these primary care physicians, over 95% listed Family Practice or Internal Medicine as their specialty.

to create one cluster for every fifty physicians within the bandwidth's length of the border (rounded up). After the initial clusters were determined, clusters were divided into "good" clusters—defined as those with 5 physicians on either side of the boundary—and "bad" clusters. The physicians in bad clusters were re-assigned to the nearest good cluster. An illustration of these clusters is displayed in Appendix Figure A4, which recreates the map from Figure 1(b) with the points for each physician and different clusters denoted by a shape-color combination.

This procedure results in the sample of providers that I use for my main analysis. I also repeat the procedure in Steps #3 and #4 for all multiples of bandwidth used in my Section 5.2 bandwidth analysis. Finally, I assign controls to each physician based on characteristics of the surrounding Census block groups. Specifically, I calculate an inverse-distance weighted average of characteristics using block groups with centroids within two miles of a given physician's location that are also in the same payment locality.

C Additional Figures and Tables

A description of the analyses performed in this section are as follows:

- **Figure A1**. This figure displays other measures of bunching to supplement the analysis in Section 4.2. See text of that section and figure notes for more detail.

- **Figure A2**. In addition to the analysis displayed in Figure 4 and discussed in Section 4.2, I also show how specific characteristics vary across the cutoff, both on the physician- and block-group-level. Panels A and C display the physician- and block-group-level results for the High-Impact Sample, respectively: of the 22 outcomes examined, only 1 achieves traditional levels of statistically significance, which is roughly the rate expected for incorrectly rejecting the null hypothesis. To emphasize that this single significant result is likely spurious and not a result of intentional sorting, I also present results for a "Low-Impact" Sample, i.e. the borders that have GAF differences of less than 2 percent. As displayed in Panels B and D, there are several characteristics that achieve statistical significance, despite the substantially lower degree of treatment (and thus reduced incentives for physicians to sort around the cutoff).

- **Figure A3**. This figure recreates Figure 5 when no adjustments for border-fixed effects are included and achieves qualitatively similar (though slightly larger) effects.

- **Figure A4**. This figure recreates Figure 1B, with overlaid points representing the location of physicians. Each color-shape combination represents a different physician cluster—see Appendix Section B and figure notes for more detail.

- **Figure A5**. This figure displays results that are analogous to those displayed in Figure 6B, but where $Y_{ip} = 1(g < GroupSize \leq g')$, where g and g' represent points the distribution of group sizes (i.e. these points represent the effects on the PDF, rather than CDF).

- **Figure A6**. This figure displays regression discontinuity plots for other measures of group size discussed in Section 5.1. See that section and figure notes for more detail.

- **Figure A7**. This figure considers two other ways that physician *treatment behavior* may change in response to differential payment. The first row (Panels A and B) represents per-patient utilization (i.e. the intensive-margin treatment decision) and the second row (Panels C and D) display the number of Medicare patients treated per physician (i.e. the extensive-margin treatment decision). Due to the imprecision of the estimates in the left-hand columns, additional controls were included in the right-hand specifications to reduce noise. (Specifically, a control for the *predicted* level of the outcome, where the prediction is generated from a LASSO design.) While the intensive-margin results do not achieve traditional levels of statistical significance ($p = 0.115$ in my most precise specification), the estimated effect of a 1.7-1.8 percent increase in resource intensity (measured by log relative value units per patient) in response

to a 1% change in reimbursement is generally in-line with recent literature on supply-side response to financial incentives (Clemens and Gottlieb, 2014). Likewise, I do not find any statistically significant effects on extensive-margin treatment decisions (measured by log Medicare patients treated), consistent with the literature on the subject.[4] Nonetheless, the 95% confidence interval for my precise specification rules out responses larger than a 1.1% increase in Medicare beneficiaries treated in response to a 1% change in reimbursement. To frame this in terms of potential *reductions* in Medicare prices, my result suggests that modest decreases in reimbursement would not limit primary care access for Medicare beneficiaries.

- **Figure A8**. This figure displays regression discontinuity plots for two other measures of practice structure, practice *location* size (Panel A) and log group size (Panel B). See Section 5.1 for further discussion.

- **Figure A9**. This figure exhibits sensitivity of estimates to the effects of individual borders by performing a "leave-one-out analysis" — i.e. re-estimating Equations 10 and 11 with all observations assigned to the listed border omitted from the sample. A shown in the figure, no individual border seems to exhibit substantial influence on the estimates.

- **Figure A10**. The purpose of this figure is to perform additional robustness checks of my main analysis:

  - "Cluster level" section: explores alternatives to the tract-level clustering strategy in the main specification.
  - "Kernel choice" section: details alternative kernel choices to the rectangular kernel used.
  - "Distance function" section: examines a different functional form for distance controls (linear was used in the main specification).
  - "Drop medical centers" line: tests robustness to omission of areas of significant provider density. While medical centers are not explicitly indicated in the data, I proxy for them by tabulating the number of providers in grids that are roughly one-third square mile in area (1/100th of a degree of longitude and latitude). I then omit observations from the analysis at the 99th percentile of provider-density among the grids.
  - "Drop phys. assigned to >1 borders" line: physicians that were within one bandwidth-distance of multiple borders were assigned to more than one locality-border (roughly 4% of my sample meets this criteria). This line tests robustness to that sample construction choice by dropping these providers.
  - "Drop phys. w/ >1 affiliation" line: roughly 10% of doctors in my sample report more than one group affiliation. As part of sample construction, I assigned the smallest listed

[4] Clemens and Gottlieb (2014) also are unable to detect any change in the number of patients treated in their setting. In contrast, Alexander and Schnell (2019) argue that an increase in *Medicaid* reimbursement for primary care services meaningfully increased healthcare access for Medicaid beneficiaries. However, because Medicaid reimbursements are substantially less generous than those for Medicare (Zuckerman et al., 2017), estimates of access from this paper are most relevant for fee structures that meet or exceed Medicare's generosity.

group size to these physicians; this tests robustness to that choice by dropping providers with multiple affiliations.

- "Drop zip-centroid geocodes" line: certain providers could only be geocoded based on the centroid of their associated zip code (rather than their exact address). Given that these providers are more likely to have a mis-measured location, they are dropped as a robustness check.
- "Specialty group FE" line: in order to more finely control for differences in physician type, I add fixed effects for each major group: Family Practice, Internal Medicine, and Other Primary Care (the Other Primary Care category covers roughly 4% of physicians in my sample).
- "Border-specific spatial controls" line: adds more detailed controls for distance by allowing linear distance controls to vary by each border.

As displayed in the figure, none of these robustness procedures changed the qualitative nature of my estimates and, in most cases, resulted only in extremely small changes to the coefficient and/or standard errors.

- **Figure A11**. This figure recreates Figure 5 when the full sample of borders are included. See Section 5.3 for more detail.
- **Figure A12**. This figure recreates Figure 6 (Panel A) and Appendix Figure A5 (Panel B) when the full sample of borders are included and the RD-Border Estimation (Equations 12 and 13) are used. See Sections 5.1 and 5.3 for more detail.
- **Figure A13.** The purpose of this figure is display heterogeneity in physician characteristics through a series of separate equations:
  - The first row displays heterogeneity by physician specialty. Given that internal medicine practice physicians generally have a larger share of Medicare beneficiaries (roughly 49% more per the 2010-2015 National Ambulatory Medical Care Surveys), it is expected that they would exhibit a stronger response to differential reimbursement. Panels A and B of the figure confirm this: the RD-IV point estimate is approximately 61% larger for internal medicine physicians than for family practice doctors (although it should be noted that these estimates are not statistically different at traditional levels).
  - The second row divides the sample by whether the physician attended a ranked medical school or not. There do not appear to be meaningful differences in these estimates.
  - The third row divides the sample by whether the physician has over/under twenty years of experience (proxied by the years between 2015 and their medical school graduation). While the RD-IV point estimates are slightly different, they are not statistically distinguishable. (It should also be noted that experience could simply be a proxy for gender

or vice versa: the male physicians comprised 69% of those with 20+ years of experience, while female physicians represented 55% of those with fewer than 20 years. This is consistent with the results of the next row.)

  - The fourth and final row divides the sample by physician gender. Again, while the RD-IV point estimates are slightly different, they are not statistically distinguishable.

- **Figure A14**. The purpose of this figure is display how well group size measures can be used to predict log fixed travel HHI ("FTHHI"). See text of Section 6 and figure notes for more detail.

- **Table A1**. This table reports all payment localities that receive different physician fee schedule payments than their respective states. See Figure 1(a) for more information.

- **Table A2**. This table reports coefficients for other physician practice style outcomes for different samples and instrumental variables. See discussion previous discussion of Appendix Figure A7 in this section for more detail regarding outcomes.

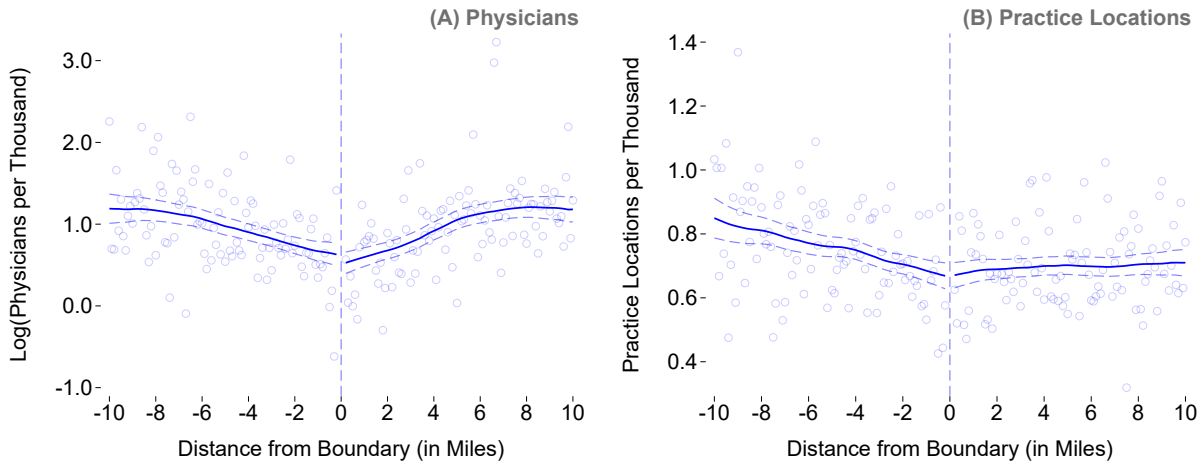
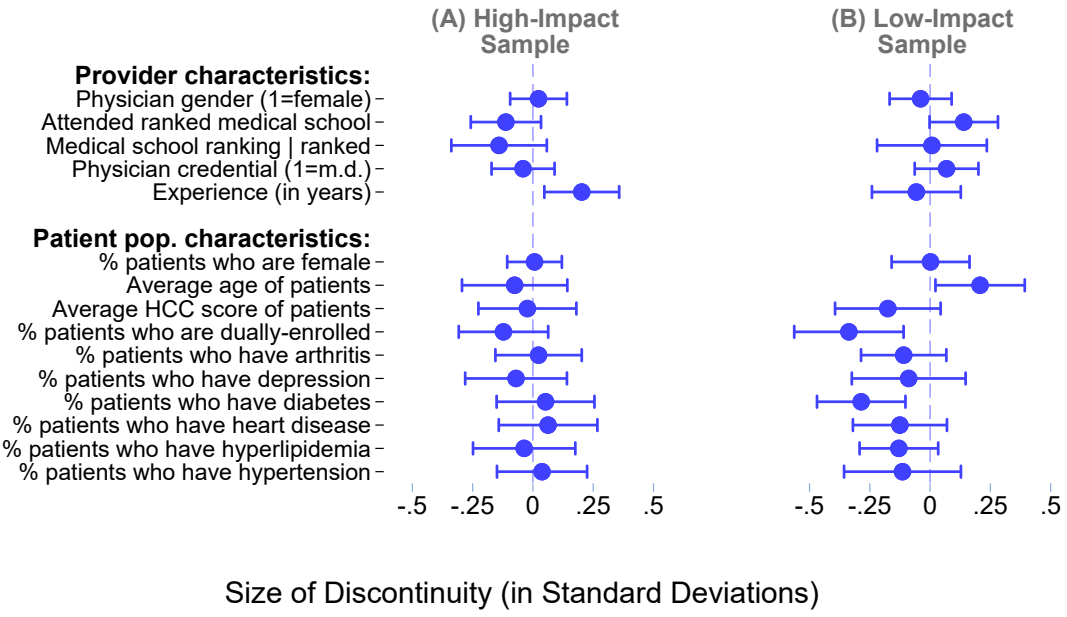


Figure A1: Validation of Research Design: Alternate Measures of Spatial Provider Density

Notes: The purpose of this figure is to illustrate the absence of bunching around locality boundaries using alternate measures to those presented in Figure 3. Panel A is analogous to Figure 3B, except with *log* physicians per thousand as an outcome measure. Panel B similarly replaces the outcome so that each point is a practice *location*, rather than an individual physician. This analysis utilizes the High-Impact Sample discussed in Section 2.2.

Source: Author calculations using CMS Physician Fee Schedule and CMS Physician and Other Supplier Public Use File.



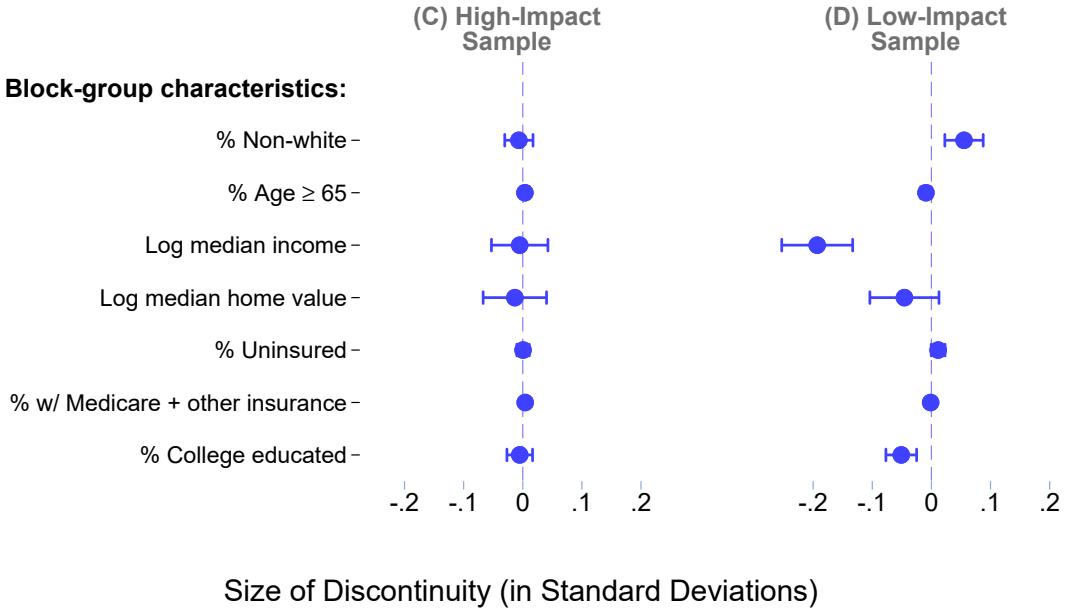


Figure A2: Validation of Research Design: Covariate Balance Tests

Notes: The purpose of this figure is to illustrate the response of selected covariates to boundary differences in Medicare reimbursement. The top panels (A and B) display provider-level regressions for provider and treated-patient-population-level characteristics, for the High-Impact Sample (borders with differences of at least two percent points) and the Low-Impact Sample (borders with differences less than two percent points), respectively. The bottom panels (C and D) display Census-block-group level regressions for local characteristics for High- and Low-Impact Samples, respectively. All variables have been normalized and used as an outcome measure in Equation 8 (i.e., these are discontinuities that have *not* been scaled by the related payment discontinuity). Bars displayed represent 90% confidence intervals. See Appendix Section C for further discussion of this figure.

Source: Author calculations using CMS Physician Fee Schedule, CMS Physician Compare Data, CMS Physician and Other Supplier Public Use File, Census Block Group 5-year ACS Estimates (Ruggles et al., 2019).

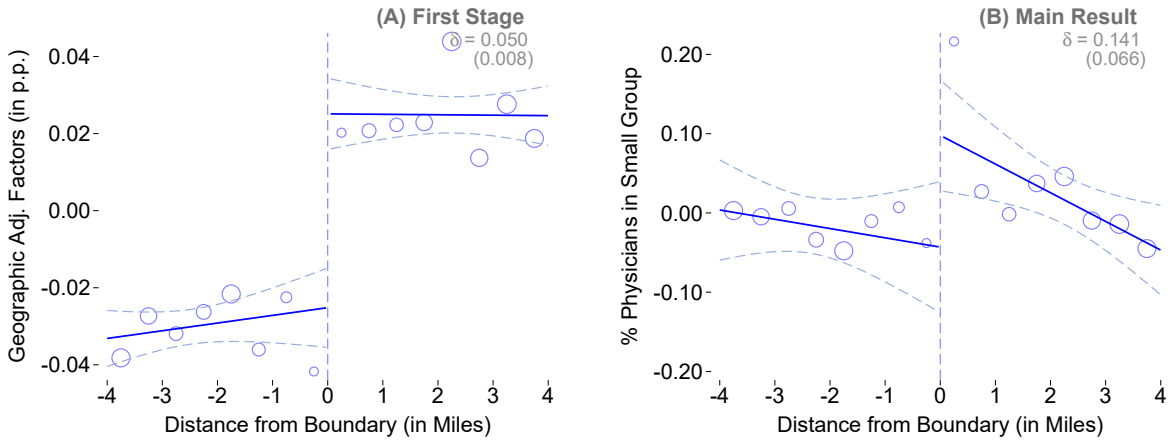


Figure A3: Main Results: Without Border FE

Notes: The purpose of this figure is to recreate the analysis displayed in Figure 5 when no adjustments for border fixed effects are included. See notes to Figure 5 for more detail.

Sources: Author calculations using CMS Physician Fee Schedule, CMS Physician Compare Data, CMS Physician and Other Supplier Public Use File.

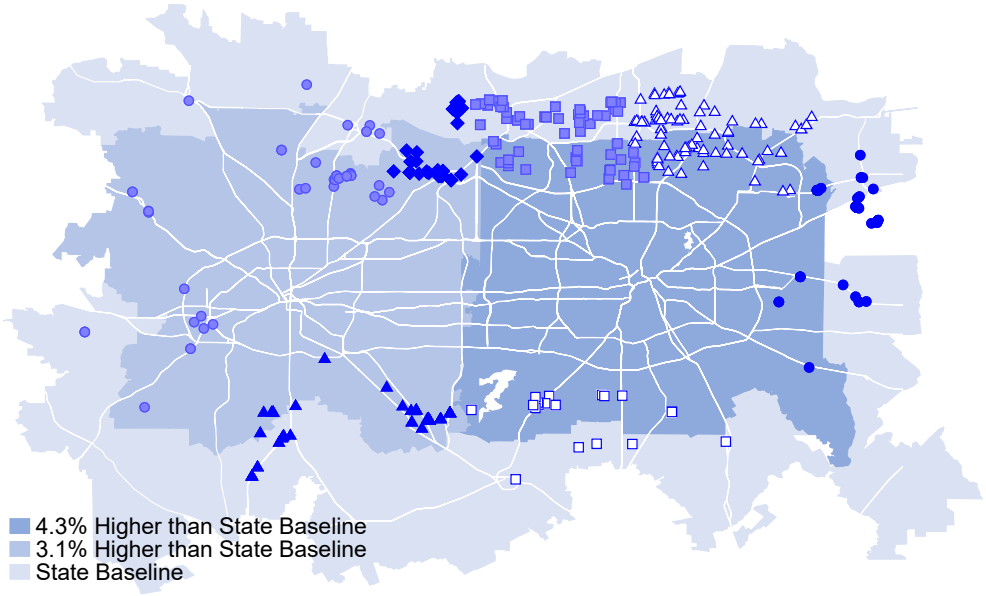


Figure A4: Localities in the Dallas-Fort Worth Area (Physician Clusters)

Notes: The purpose of this figure is to illustrate the physician clusters used analyses employing cluster-by-border fixed effects (discussed further in Section 4). Each shape-by-color combination represents a different set of physicians that are grouped together. Use of border-by-cluster fixed effects effectively forces comparisons within these physician groups. See notes to Figure 1(b) for further description of the underlying map characteristics. See notes to Appendix Section B for more detail on how these clusters were constructed.

Sources: Author calculations using CMS Physician Fee Schedule, CMS Physician and Other Supplier Public Use File.

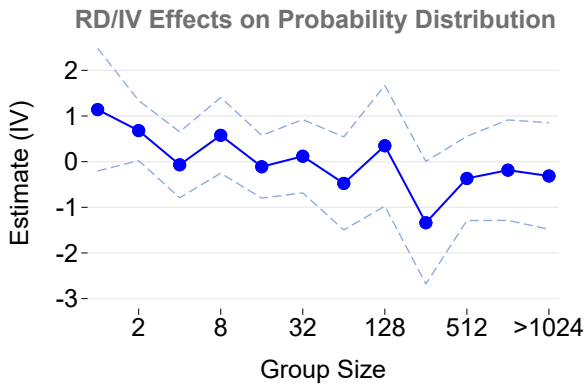


Figure A5: Distributional Effects on Group Size

Notes: The purpose of this figure is to display results that are analogous to those displayed in Figure 6B, but where $Y_{ip} = 1(g < GroupSize \leq g')$, where g and g' are points in the distribution of group sizes. Coefficient estimated, presented along with 90% confidence intervals, show the impact of a 1 percent difference in Medicare reimbursement is associated with a δ p.p. change in the displayed probability.

Source: Author calculations using CMS Physician Fee Schedule, CMS Physician Compare Data, and CMS Physician and Other Supplier Public Use File.

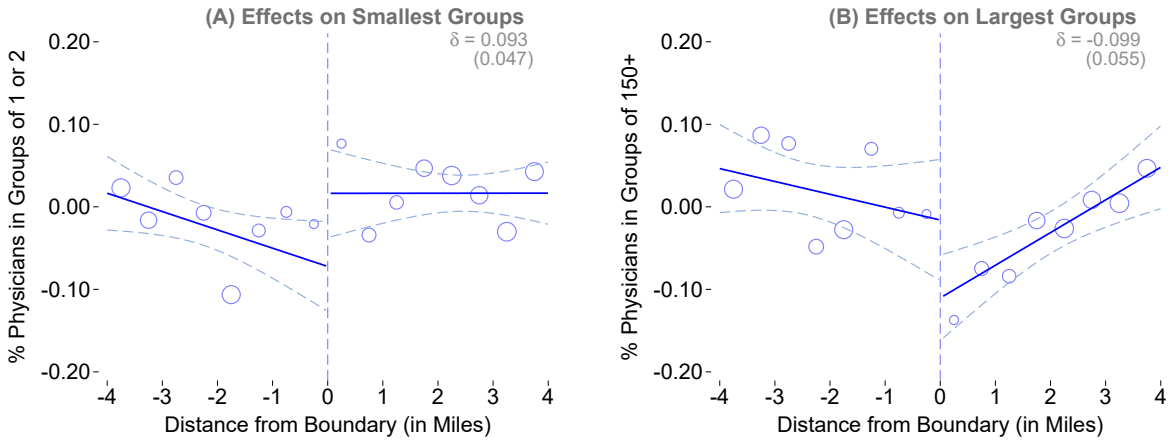


Figure A6: Distributional Results: Effects on Smallest and Largest Groups

Notes: The purpose of this figure to is present regression discontinuity plots analogous to Figure 5 for two other measures of group size. Panel A represents the impact of differential payments on the smallest groups (1-2 providers) while Panel B displays the results for the largest groups (150 providers or more). See Figure 5 for more detail on the structure of these plots.

Sources: Author calculations using CMS Physician Fee Schedule, CMS Physician Compare Data, CMS Physician and Other Supplier Public Use File.

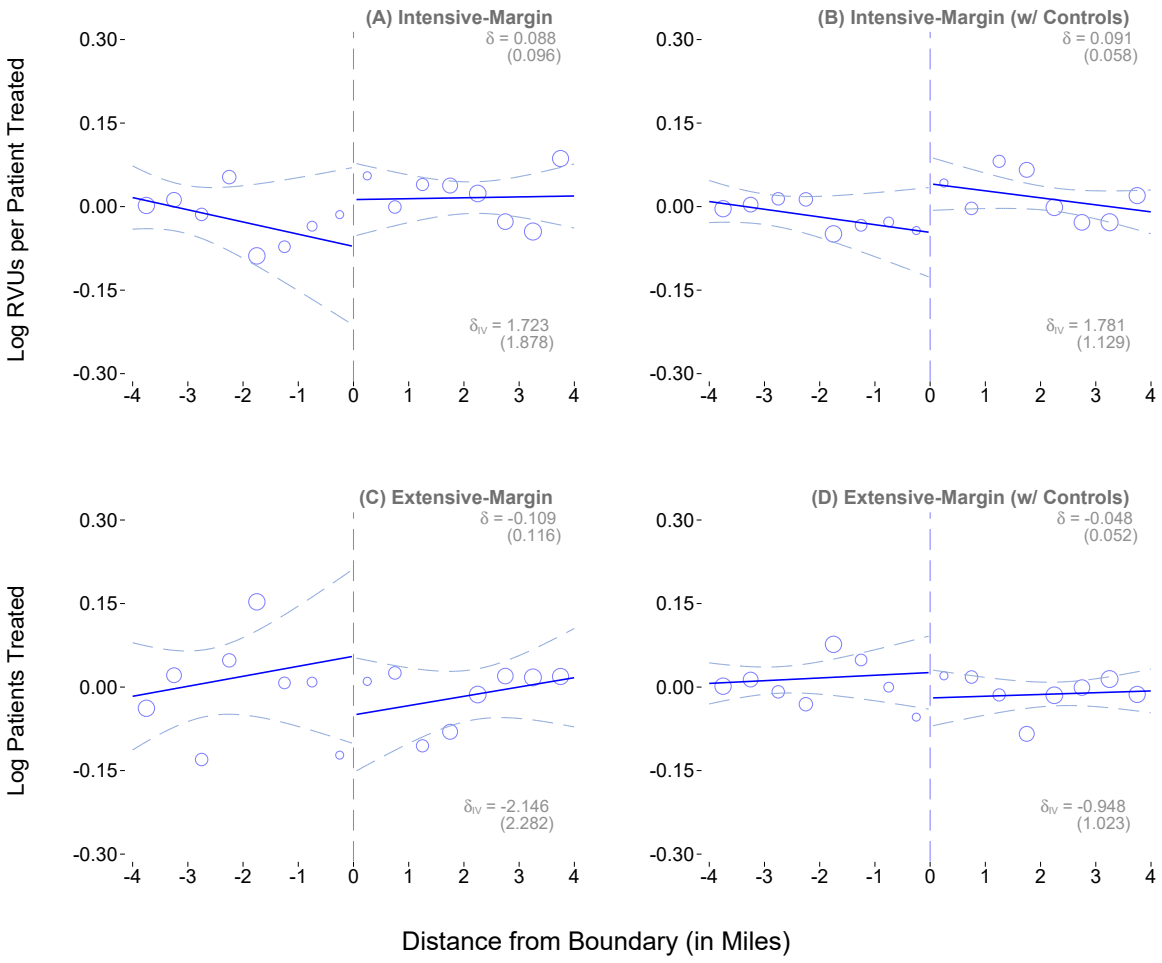


Figure A7: Other Results: Treatment Outcomes

Notes: The purpose of this figure to is present regression discontinuity plots analogous to Figure 5 for measures of physician practice style. See Appendix Section C for further discussion and Figure 5 for more detail on the structure of these plots.

Sources: Author calculations using CMS Physician Fee Schedule, CMS Physician Compare Data, CMS Physician and Other Supplier Public Use File.

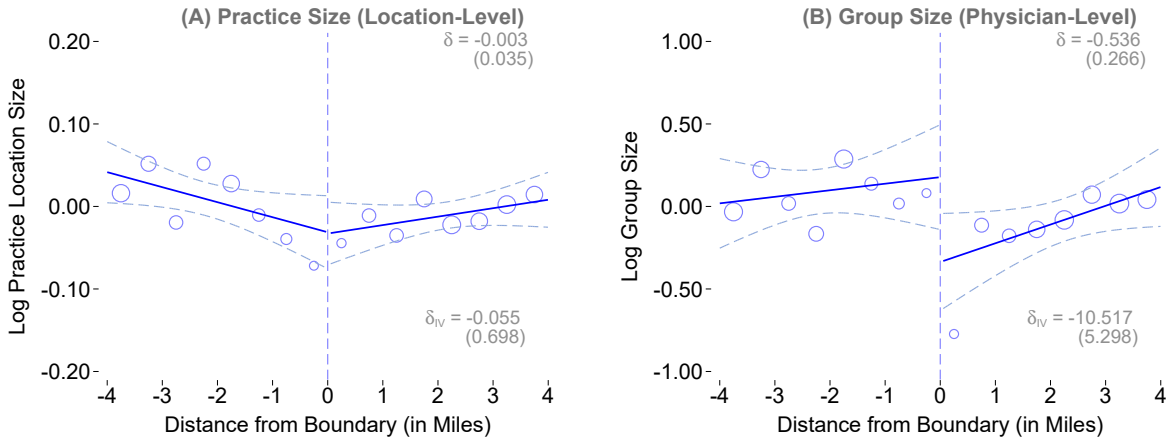


Figure A8: Other Results: Other Practice Structure Outcomes

Notes: The purpose of this figure to is present regression discontinuity plots analogous to Figure 5 for other measures of practice structure. See Section 5.1 for further discussion and Figure 5 for more detail on the structure of these plots.

Sources: Author calculations using CMS Physician Fee Schedule, CMS Physician Compare Data, CMS Physician and Other Supplier Public Use File.

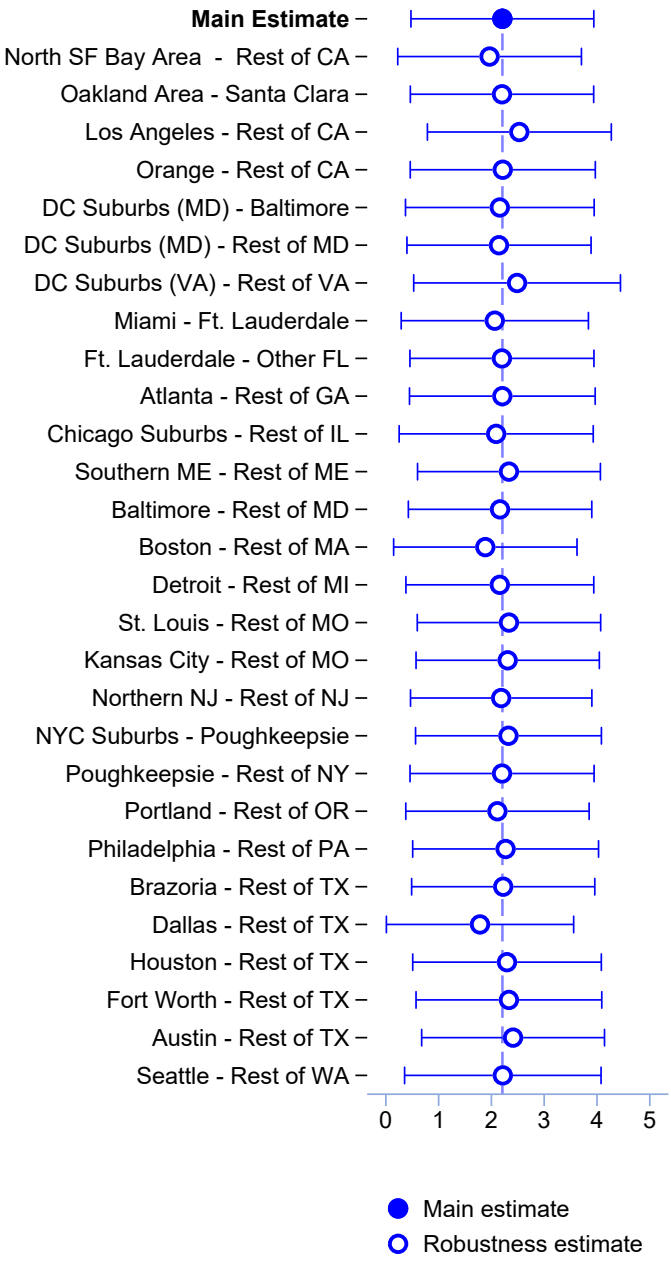


Figure A9: Border Leave-one-out Analysis

Notes: This figure exhibits sensitivity of estimates to the effects of individual borders. Accordingly, each coefficient presented represents a separate regression of the preferred specification of Equations 10 and 11 with the listed border name omitted from the sample. Bars indicate 90% confidence intervals.

Sources: Author calculations using CMS Physician Fee Schedule, CMS Physician Compare Data, CMS Physician and Other Supplier Public Use File, Census Block Group 5-year ACS Estimates (Ruggles et al., 2019).

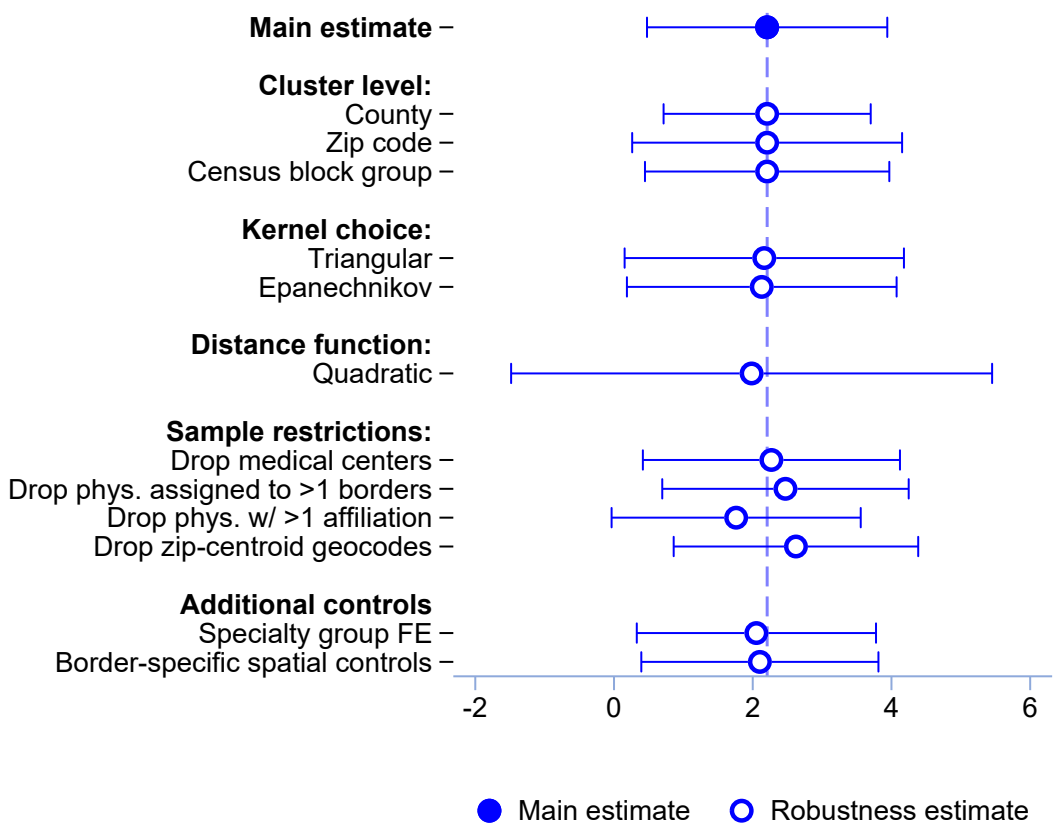


Figure A10: Other Robustness Checks

Notes: The purpose of this figure is to perform additional robustness checks of my main analysis. Each point represents a separate regression of the preferred specification of Equations 10 and 11 with the listed modification. Bars indicate 90% confidence intervals. See the text of Appendix Section C for description of each robustness test.

Sources: Author calculations using CMS Physician Fee Schedule, CMS Physician Compare Data, CMS Physician and Other Supplier Public Use File, Census Block Group 5-year ACS Estimates (Ruggles et al., 2019).

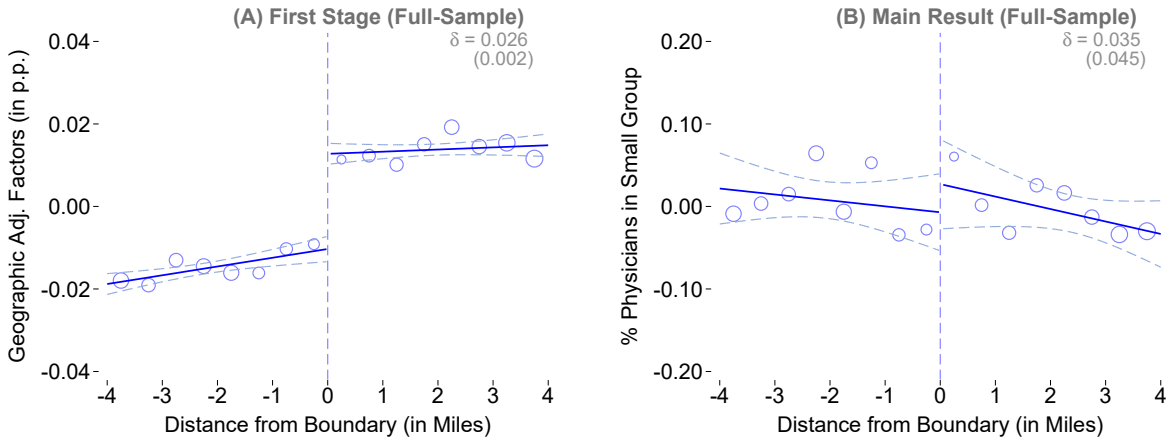


Figure A11: Main Results: Full Sample

Notes: The purpose of this figure is to recreate the analysis displayed in Figure 5 when all borders are included. See discussion in Section 5.3 and notes to Figure 5 for more detail.

Sources: Author calculations using CMS Physician Fee Schedule, CMS Physician Compare Data, CMS Physician and Other Supplier Public Use File.

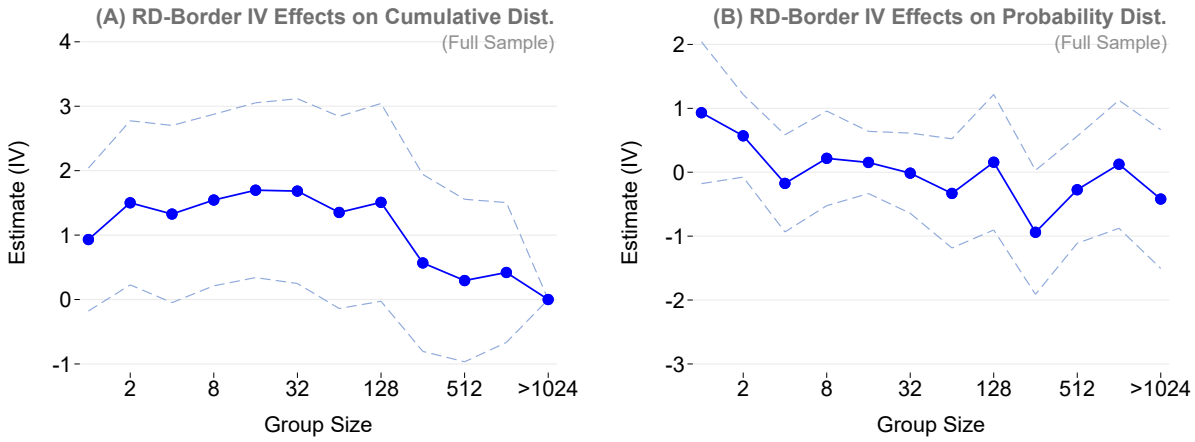


Figure A12: Distributional Effects on Group Size

Notes: The purpose of this figure is to recreate Figure 6 (Panel A) and Appendix Figure A5 (Panel B) when the full sample of borders are included and the RD-Border Estimation (Equations 12 and 13) are used. See Sections 5.1 and 5.3 and notes to referenced figures for more detail.

Source: Author calculations using CMS Physician Fee Schedule, CMS Physician Compare Data, and CMS Physician and Other Supplier Public Use File.

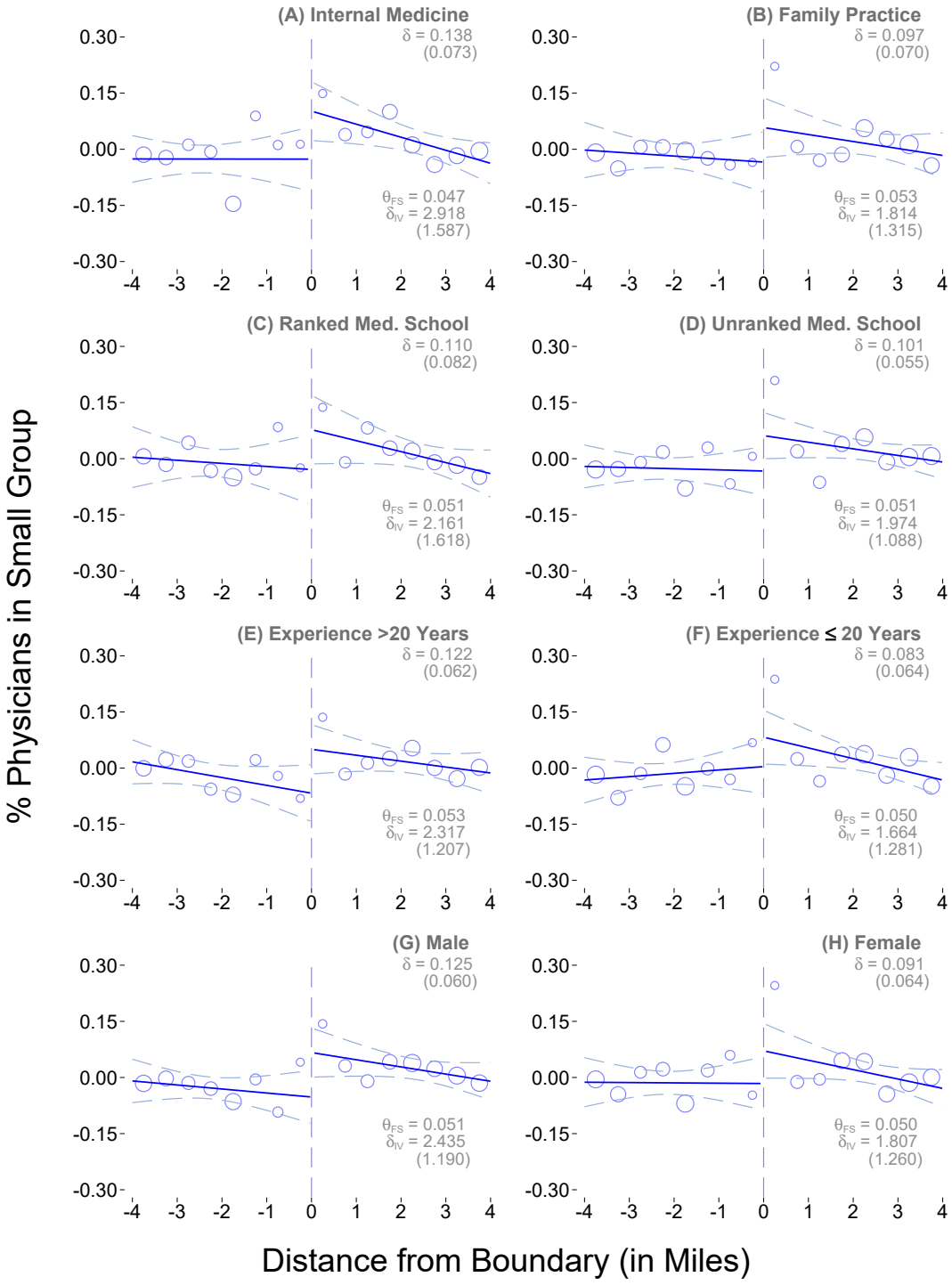


Figure A13: Heterogeneity by Physician Characteristics

Notes: The purpose of this figure is display heterogeneity of results by physician characteristics. Within each panel, reduced-form estimates (Equation 8) and associated standard errors are in the upper-right corner, while the first-stage (Equation 11) and RD-IV estimate (δ_{IV}, from Equation 10) and associated standard errors are in the bottom-right corner. See discussion in Appendix Section C for more detail.

Sources: Author calculations using CMS Physician Fee Schedule, CMS Physician Compare Data, CMS Physician and Other Supplier Public Use File.

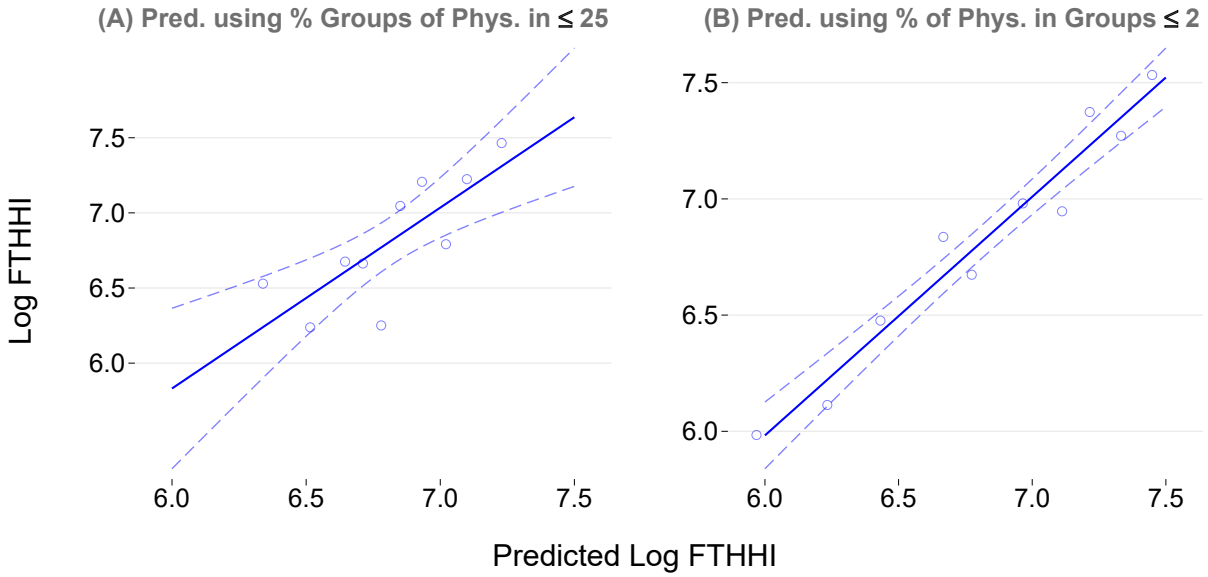


Figure A14: Prediction of Log FTHHI by Group Size Measure

Notes: The purpose of this figure is display how well group size measures can be used to predict log fixed travel HHI ("FTHHI"). Panel A utilizes the main outcome, % of providers in a small group (defined as 25 providers or fewer), while Panel B utilizes the best predictor among the statistically significant distributional results, the % of providers in groups of 1-2. Each point of the graph represents a binned decile of predicted log FTHHI.

Sources: Author calculations using CMS Physician Fee Schedule, CMS Physician Compare Data, CMS Physician and Other Supplier Public Use File, Census Tract 5-year ACS Estimates (Ruggles et al., 2019).

Table A1: Medicare-Designated Payment Localities

State	Metro Area or County	State	Metro Area or County	State	Metro Area or County
California	Marin/Napa/Solano	Illinois	Suburban Chicago	New York	Queens
California	San Francisco	Illinois	Chicago	Oregon	Portland
California	San Mateo	Louisiana	New Orleans	Pennsylvania	Metro Philadelphia
California	Oakland/Berkley	Maine	Southern Maine	Texas	Brazoria
California	Santa Clara	Maryland	Baltimore/Surr. Ctys	Texas	Dallas
California	Ventura	Massachusetts	Metro Boston	Texas	Galveston
California	Los Angeles	Michigan	Detroit	Texas	Houston
California	Anaheim/Santa Ana	Missouri	Metro St. Louis	Texas	Beaumont
District Of Columbia	DC + MD/VA Suburbs	Missouri	Metro Kansas City	Texas	Fort Worth
Florida	Fort Lauderdale	New Jersey	Northern NJ	Texas	Austin
Florida	Miami	New York	Manhattan	Washington	Seattle (King County)
Georgia	Atlanta	New York	NYC Suburbs		
Illinois	East St. Louis	New York	North NYC Suburbs		

Notes: This table reports all payment localities that receive different physician fee schedule payments than their respective states. See Figure 1(a) for more information. Note that some areas listed here are not included in Figure A9 as certain borders did not have any points meeting the criteria discussed in Section B.

Sources: CMS Physician Fee Schedule.

Table A2: Alternate Results: Treatment Outcomes

Outcomes	**(1)**	**(2)**	**(3)**	**(4)**
Log RVUs per patient	1.723	1.894	1.288	1.242
	(1.878)	(2.250)	(1.438)	(1.421)
Log unique patients	-2.146	1.117	-1.062	-0.891
	(2.282)	(2.836)	(1.829)	(1.798)
Sample	High Impact	Full	High Impact	Full
Instrumental variable	RD-IV	RD-IV	RD-Border	RD-Border
Observations	5,065	9,685	5,065	9,685
First stage F-statistic	399	148	7,468	5,920

Notes: This table recreates the analysis in Table 2 for physician practice style variables. See Section C for further discussion of these variables and Section 5.3 for further discussion of alternate samples and instrumental variables. See also Table 2 for more information on table structure.

Source: Author calculations using CMS Physician Fee Schedule, CMS Physician Compare Data, CMS Physician and Other Supplier Public Use File.