NBER WORKING PAPER SERIES

INFORMATION AND DISPARITIES IN HEALTH CARE QUALITY: EVIDENCE FROM GP CHOICE IN ENGLAND

Zach Y. Brown
Christopher Hansman
Jordan Keener
Andre F. Veiga

Working Paper 31033
http://www.nber.org/papers/w31033

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2023

We are grateful to Martin Gaynor, Matt Grennan, Ben Handel, Maria Polyakova, Carol Propper, and seminar participants at APPAM, Arizona, Berkeley, Caltech, the CEPR virtual IO seminar, Duke, Groningen, GSU, Johns Hopkins, Maryland, Michigan, NBER Summer Institute, SITE, Toulouse, Tsinghua, UC Irvine, UGA, WBEC, and the ZEW Conference on Information and Communication Technologies for valuable comments. We thank Radoslaw Kowalski for sharing data. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Information and Disparities in Health Care Quality: Evidence from GP Choice in England
Zach Y. Brown, Christopher Hansman, Jordan Keener, and Andre F. Veiga
NBER Working Paper No. 31033
March 2023
JEL No. D83,I11,I14,L15

ABSTRACT

Why do low income patients tend to go to lower quality health care providers, even when they are free? We show that differential information about provider quality is an important determinant of this disparity. Our empirical strategy exploits the temporary presence of a website that publicly displayed summary star ratings of general practitioner (GP) offices in England. Regression discontinuity estimates show that patient demand responds sharply to the information on the website, and that this response is almost entirely driven by residents of low income neighborhoods. The results are consistent with high income patients having better private information about quality. We incorporate our estimates into a structural model of demand that allows for heterogeneity in information, preferences, and consumer inertia. We find that information differences explain 24 percent of the relationship between income and GP quality and reinforce disparities in access to care.

Zach Y. Brown
Department of Economics
University of Michigan
611 Tappan Ave
Ann Arbor, MI 48109
and NBER
zachb@umich.edu

Christopher Hansman
Imperial College Business School
Imperial College London
South Kensington Campus
London SW7 2AZ
UK
c.hansman@imperial.ac.uk

Jordan Keener
University of Michigan
Department of Economics
jordkeen@umich.edu

Andre F. Veiga
Imperial College London
a.veiga@imperial.ac.uk

1 Introduction

There is persistent within-country inequality in health care throughout the world (Cookson et al. 2021; Hart 1971). Even in countries with free public health services, like the United Kingdom, lower-income individuals receive lower quality care (van Doorslaer et al. 2004, 2006; OECD 2014; Cookson et al. 2016; Scobie and Morris 2020).[1] While policymakers often fixate on logistical barriers to *access*, inequality in information—for example, because some patients are better able to research providers or obtain recommendations from social networks—may also drive these disparities.[2] Disentangling the roles of information and access is challenging, particularly if patient preferences also vary across the income distribution. However, understanding the root causes of differences in care is crucial for the design of an equitable and efficient health system.

In England, General Practitioners (GPs) are the first point of contact with the health care system, and there are large disparities in quality over the income distribution. As we show, individuals in high income neighborhoods register with better GP practices across a wide range of metrics.[3] A portion of this can be explained by access because providers located in more affluent areas are typically of higher quality. However, even fixing the set of choices, low income patients typically select lower-quality options. This correlation could be a result of differences in information or could reflect differences in preferences that affect how individuals trade-off quality, travel distance, and other GP characteristics.

Our central contribution is to show that information is a key driver of these disparities. To do so, we combine regression discontinuity evidence with a rich structural model of patient choice. Our approach revolves around the presence of a widely used patient review system shown on the National Health Service (NHS) website. Until 2020, the website provided a summary star rating for each GP, measured as the average patient review over the past two years, rounded to the nearest half star. This rounding creates a series of discontinuities in the distribution of average reviews, providing sharp variation in public information about quality for very similar providers.

[1] Similar patterns are present in the United States (AHRQ 2021).

[2] The National Health Service in England emphasizes barriers to access, noting that it undertakes "considerable work to meet its moral and legal obligations to promote equality and address health inequalities to improve access to services" (NHS 2016).

[3] This pattern, and the demand for quality in general, is consistent with previous work (Santos et al. 2017).

To show why the rating system helps us isolate the role of information, we present a theoretical framework in which patients combine private information with a publicly observed rounded star rating to form beliefs about quality. The model delivers direct predictions about the relationship between ratings, private information, and patient choice. The simple take away is that privately well informed individuals should not respond much to the website's star ratings. Their beliefs will be closely tied to provider quality overall, but only marginally influenced by the coarse public signal provided by the website. On the other hand, less informed patients will be strongly influenced by star ratings but relatively insensitive to quality conditional on the information provided by the website.

Motivated by this framework, we turn to our regression discontinuity (RD) approach, which compares patient demand for GPs just above versus below the rounding thresholds for star ratings.[4] We find sharp jumps in demand at each threshold, and show that this effect is almost entirely driven by low income patients. In the period after 2020, when the star ratings were removed, the jumps disappear. We do not find any corresponding discontinuities in wait times or other observable characteristics of GPs and no bunching in the distribution of GPs. We find similar results for the subset of individuals who choose a new GP because they change address, and with a panel-regression strategy focusing on GPs that experience star rating changes over time.

Our RD results are consistent with disparities in information between low and high income patients. We find that low income patients respond sharply to star ratings—even for GPs with nearly identical quality—but are not sensitive to differences between GPs with the same number of stars. Alternatively, high income patients respond to small differences in quality, but are not meaningfully influenced by the star ratings themselves. Through the lens of our model, these results suggest that high income patients have substantially more precise private information about GP quality.

To quantify the role of information gaps relative to other drivers of inequality, we develop and estimate a structural model of GP choice based on our theoretical framework. The model incorporates (and allows for heterogeneity by income in) uncertainty over provider quality. To capture the fact that patients switch GPs infrequently, we allow for consumer inertia.

[4] E.g., comparing practices rated 3.74 which round down to 3.5 versus 3.76 which round up to 4. This follows previous literature using star rating systems, such as Anderson and Magruder (2012) and Luca (2016).

The model also includes GP fixed effects, to capture unobservable dimensions of GP quality, allows income heterogeneity in preferences for quality and travel, and accounts for patient crowding that could make otherwise high quality GPs unappealing due to wait times.

We estimate the model using an indirect inference approach (Gourieroux et al. 1993) that incorporates our RD results. Targeting the RD estimates in the estimation procedure helps us disentangle heterogeneous information and heterogeneous preferences as drivers of GP demand. In addition, we target moments constructed from GP market shares and patient switching rates. In line with our reduced form results, estimates from the demand model imply that high income individuals have more precise information about quality. We do not find substantial differences in preferences for quality, although high income individuals have slightly less distaste for travel distance. The model matches our targeted moments closely and can rationalize key patterns in the data.

Our structural approach allows us to evaluate the drivers of health care disparities by considering counterfactual simulations in which we eliminate differences in information or access. Relative to the post-2019 status quo (without the public star rating system), equalizing the information of high and low income patient removes a substantial fraction of the relationship between income and quality, on the order of 24%. Alternatively, removing barriers to access by equating choice sets across individuals reduces disparities by 55%. Almost 90% of the relationship disappears if we equalize information and access to quality GPs simultaneously, suggesting that information and access are complements. Differences in preferences between low and high income patients are not a major driver of disparities.

Our results indicate that relatively low cost informational interventions can be effective in reducing health care inequality. The star rating system is one example—and we show that this coarse information was partially able to address information disparities—but other interventions may be more effective. Additionally, our results suggest that increasing accessibility alone, while important, will not be sufficient to eliminate health care inequalities. A combination of information and access is necessary, and there are synergies when pairing the two. Put simply, increasing access will not benefit patients who are unaware of or unable to identify high quality providers.

Our primary contribution comes in providing evidence on—as well as modeling and quantifying the role of—informational differences as a driver of inequality in health care choice. This serves as a bridge between two large literatures. The first is a broad body of work highlighting the existence and persistence of healthcare inequalities by income (e.g., Hart 1971; Peters et al. 2008; van Doorslaer et al. 2006; Gwatkin et al. 2004; Cookson et al. 2016; Marmot et al. 2007; Balarajan et al. 2011; Devaux 2015) as well as disparities in health insurance plan choice (Handel et al. 2020). The second is a literature documenting the responsiveness of consumers to information about healthcare quality and associated equilibrium effects (e.g., Dranove et al. 2003; Cutler et al. 2004; Pope 2009; Kolstad 2013. For a review, see Kolstad and Chernew 2009). A related literature finds evidence of consumer uncertainty about quality in other health care settings (e.g., Jin and Sorensen 2006; Dafny and Dranove 2008; Werner et al. 2012; Darden and McCarthy 2015; Grennan and Town 2020). There is also evidence that physicians and their families make different health care decisions, including drug choice and preventative care, consistent with having more health care expertise (Bronnenberg et al. 2015; Chen et al. 2019; Artmann et al. 2022). Our innovation comes in highlighting information as an important source of disparities in the choice of health care provider.

More generally, the mechanism we highlight echos similar evidence in other contexts, perhaps most notably in education. For example, Hastings and Weinstein (2008) show that the provision of information about quality impacts school choice for low income households, while Kapor et al. (2020) find that families of students in poor neighborhoods have more dispersed beliefs about admissions probabilities (see also Bettinger et al. 2012; Hastings et al. 2015). We document similar channels in the health care context, where inequality is a first order concern and there is substantial evidence of heterogeneity in quality (Doyle et al. 2019; Hull 2018; Cooper et al. 2022).

Finally, our paper relates to a growing body of work that analyzes online rating systems, including Chevalier and Mayzlin (2006), Anderson and Magruder (2012), Lewis and Zervas (2016), Luca (2016), Luca and Zervas (2016), Reimers and Waldfogel (2021) and more, particularly those that explicitly incorporate learning models (e.g., Xiao 2010; Newberry and Zhou 2019). Recent work has examined the effect of review systems on doctor congestion and interruptions in care (Chartock 2021; Kummer et al. 2021). Our primary contribution is in using the response to reviews to identify information disparities.

The remainder of this paper is as follows. Section 2 provides additional background on the website and data. Section 3 presents a stylized model of GP choice in the presence of uncertainty and a rating website. Section 4 introduces and present results from our RD design. Section 5 presents our structural model, estimates, and counterfactual simulations. Section 6 concludes.

2 Background, Data and Disparities in GP Quality

In this section we provide background on how individuals choose a general practitioner (GP) in England, including details on the government website providing information about GP practices. We then outline the data used in our analyses and provide descriptive evidence on the existence of disparities in GP quality by income.

2.1 Background on GPs and the NHS Choice Website

Choosing a GP practice

In England, GPs are the first point of contact for patients. GPs provide a wide range of services, including checkups, screenings, vaccinations, simple surgeries, and referrals to secondary care. They are organized into practices comprising several doctors, which contract directly with the National Health Service (NHS).[5]

All residents of England can register with a GP practice free of charge. Individuals register with a practice, not an individual doctor (for simplicity, we refer to GP practices, potentially comprising multiple doctors, as "GPs"). Absent capacity constraints or a set of specialized circumstances, practices must accept all patients who wish to register.[6] The process of enrollment is straightforward and can often be completed online. Patients are free to switch GPs at any time, and medical records are automatically transferred. GPs

[5]The primary funding of GPs is based on the number of patients enrolled at the practice, weighted by patient characteristics. This is supplemented with performance pay based on schemes such as the Quality and Outcomes Framework (QOF), if a GP voluntarily participates. For more details, see `https://www.bma.org.uk/advice-and-support/gp-practices/funding-and-contracts/global-sum-allocation-formula`.

[6]The right to choose a GP is directly outlined in the constitution of the NHS. The constitution states "You have the right to choose your GP practice, and to be accepted by that practice unless there are reasonable grounds to refuse." As of January 2015, patients can choose a GP outside their designated geographic area. The practice is allowed to refuse a patient if there is concern that the patient lives too far away and traveling will be inconvenient or dangerous given the patient's health status. See `https://www.gov.uk/government/publications/the-nhs-choice-framework`.

do not advertise. The practice of selectively enrolling patients is likely minimal or non-existent: there is limited heterogeneity in costs for primary care, funding is weighted by patient complexity, and GPs face legal disincentives to cream-skim patients.

There is concern that capacity constraints or wait times for appointments may impact patient choices. Our central focus is on enrollment with a GP, which is distinct from arranging an appointment. A patient may enroll (and appear in our data) even if they face significant delays in actually seeing a doctor. In fact, strict capacity constraints are quite rare. At any given time, less than 2% of GPs are officially closed to new enrollment due to having too many patients.[7] Of course, the anticipation of wait times may dissuade new patients from registering with a specific GP. To account for this, we collect data on both GP crowding and patient wait times and use this information in our analysis.

Online Reviews

The NHS maintains a website (`www.nhs.uk`) that provides information to help patients choose a GP practice, hospital, dentist, or other health care provider.[8] This website currently receives over 1 billion visits per year.[9] We focus on a rating system for GP practices included on the website that, during our main sample, was referred to as *NHS Choices*. This system allows patients to leave a written review of a GP practice and to provide a ranking from one to five "stars," similar to online review platforms like Yelp or TripAdvisor. Between 2007 and the end of 2019, these reviews were used to construct *star ratings* which were displayed prominently at the top of each GP practice's page as well as in search results. These summary star ratings were calculated as the average rating over the previous two years. Importantly, these ratings were rounded to the nearest half-star (e.g., an average rating of 3.26 stars is shown as 3.5 stars). An example of the platform for one GP is shown in Panel A of Appendix Figure A-1. The summary star rating is shown in the upper right-hand corner. Given that the website was widely known and heavily trafficked, it is reasonable to assume that most patients

[7]According to Freedom of Information requests, only 145 GPs were approved to be closed to new patients during some part of the period 2016/17 and only 106 GPs were approved to be closed to new patients in 2017/18. See FOI-056173. We are not aware of data allowing us to identify the specific GPs that do not accept new patients.

[8]The website also provides details on healthcare and pharmaceutical services alongside general purpose medical information. Some GP practices allow patients to access their records, make appointments, and order repeat prescriptions through the website.

[9]See `https://digital.nhs.uk/news/2022/1.2-billion-visits-to-nhs-website-in-last-12-months`.

searching for a GP saw these star ratings.

At the end of 2019, the website removed the summary star ratings from the provider pages but kept the individual ratings and text comments by users (see Panel B of Appendix Figure A-1). In theory, individuals could still manually calculate each GP's star ranking, but the information was significantly more difficult to access. Our primary analysis considers only the period prior to January 2020 in which the star ratings were directly displayed. However, we consider the current status quo—with no star ratings—in falsification exercises and as a benchmark for counterfactual simulations.

2.2 Data

Review and Star Ratings

We construct a monthly panel of individual reviews on the website, from May 2015 to December 2021.[10] For each GP practice, the panel contains the index r_{jt}, which is the average across all ratings of practice j in the two years preceding month t. We take this as our key measure of quality. We calculate the star rating s_{jt} as r_{jt} rounded to the nearest half star, which is what was observed by patients on the website prior to 2020.

There is a concern that an index based on user feedback may not be a meaningful measure of clinical quality. Analysis of written reviews indicates that patients judge providers based on a wide variety of concerns, including quality of the medical services, amount of bureaucratic red tape, bed-side manner, and the quality of the facilities (Kowalski 2017). We verify that average reviews capture a consequential dimension of quality by examining the correlation between the index r_{jt} and other subjective and objective measures of quality in Appendix Table A-1. For instance, we find a correlation of 0.52 with a measure of overall experience based on representative patient surveys. The index r_{jt} is also highly correlated with the NHS Quality and Outcomes Framework (QOF) clinical score, a measure of provider quality that includes objective clinical indicators.[11] All correlation coefficients are statistically significant, providing evidence that patient feedback reflects, at least in part, objective

[10] We collected individual reviews and rankings for each GP practice from the NHS Choices website for the period April 2016 to December 2021. We combine this with previously collected data for the period May 2013 to April 2016. See Kowalski (2017).

[11] We provide additional details on the patient survey and QOF scores in Appendix Section A.1.

clinical quality. Appendix Table A-1 is consistent with previous evidence documenting a correlation between patient surveys and objective health outcomes, including evidence that hospital ratings on NHS Choices correlate with mortality and readmission rates (Greaves et al. 2012b,a). However, it is clear that individuals also care about non-clinical dimensions of GP practice quality. We are a priori agnostic about how much weight individuals place on clinical dimensions of quality, and our structural model (Section 5) allows us to obtain additional insight into patient preferences over a variety of GP characteristics.

A second concern is that there may be credibility issues, particularly if fake ratings are common (as has been documented in other review systems, see, e.g. Mayzlin et al. 2014; Luca and Zervas 2016). Provider ratings on this website are government sanctioned and moderated by the NHS, which collects information on each individual leaving a review, including email and IP addresses.[12] This moderation discourages explicit gaming by providers or their employees and helps ensure that reviews are informative. The fact that ratings are highly correlated with independent measures of quality, including representative surveys, further helps assuage credibility concerns. Given our RD approach, a particular concern would be that GPs close to rounding thresholds might tamper with or submit reviews to get a more favorable star rating, something we find evidence against in Section 4.

GP Enrollment and Characteristics

We match the review and star rating data with enrollment data for the universe of GP practices in England.[13] For each GP, we observe quarterly enrollment by Lower Super Output Area (LSOA). LSOAs are detailed geographic areas with an average of about 700 households (2,000 individuals), and are roughly analogous to census block groups in the US. We construct a quarterly panel of enrollment at the GP-LSOA level. GP enrollment is highly persistent over time, since patients tend not to switch GPs unless they switch addresses. For this reason, we focus on net new enrollment in our reduced form analysis, measured as the quarterly change in enrollment at the LSOA level, as our primary outcome.

We merge on LSOA characteristics from the 2019 English Indices of Deprivation, includ-

[12]See www.nhs.uk/our-policies/comments-policy.

[13]These data were obtained from the Primary Care Registration database within the National Health Application and Infrastructure Services (NHAIS) system.

ing income, health, education, and employment. Our measure of neighborhood income is the standardized fraction of individuals that are low income.[14] We also employ quarterly data on GP characteristics such as the number of practitioners, mean experience of practitioners, and the age of the practice.[15] We geocode addresses for all GPs, allowing us to calculate the distance from each GP to the centroid of each LSOA.

We supplement our analysis with restricted individual-level enrollment records obtained via a data access request to the NHS. These data include LSOA information on all individuals that change addresses, who we label movers. In addition, the records include information on the subset of patients that join a GP in each month, including an individual identifier, age, the month of a switch, the GP that an individual joined and the patient's most recent LSOA of residence. While our primary analysis focuses on the universe of all patients in England, many individuals switch GPs only when they switch residences. The individual-level data allow us to replicate our analysis for the subset of individuals that were forced to actively choose a GP because they no longer live near their old provider. The data also allow us to construct additional moments to better account for inertia in our structural model.

2.3 Summary Statistics

The first panel of Table 1 shows summary statistics for the GP rating system. The first two columns show data from the period in which star ratings were visible (through December 2019), the second two columns show the later period. Our sample includes 7,635 unique GP practices, over 18 million GP×LSOA×quarter observations and over 350,000 individual reviews. During this period, the mean GP has an index (r_{jt}) of 3.2 stars. There is a large degree of dispersion of GP quality as measured by star ratings. The standard deviation across practices is 1.0 stars and over 10 percent of GPs have 2 stars or less. Appendix Figure A-2 shows that the distribution of average reviews r_{jt} across GP practices is relatively smooth, including near the rounding thresholds.

Table 1 also shows summary statistics on enrollment and patient demographics. The average practice has about 8,000 patients. Mean enrollment at a practice from a given LSOA

[14] For details on how income deprivation is computed, see https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019.

[15] These data were obtained from NHS Digital. We describe GP characteristics and other supplemental data sources in Appendix Section A.1.

Table 1
Summary of GP Enrollment and Characteristics

	Period with Star Ratings		Period without Star Ratings	
	Mean	SD	Mean	SD
GP Reviews:				
Individual review	3.17	1.84	3.43	1.69
GP average stars	3.20	1.02	.	.
GP Number of Reviews	84.5	89.7	122.6	146.0
GP Enrollment:				
Total Enrollment (100s)	80.75	50.92	91.38	60.50
LSOA Enrollment (100s)	0.58	1.61	0.54	1.61
Quarterly LSOA Enrollment Change	0.17	2.08	0.09	1.81
Average GP Patient Demographics:				
Female	0.50	0.02	0.50	0.10
Age	39.92	4.54	40.29	4.56
LSOA Income deprivation	0.13	0.10	0.13	0.10
LSOA Health deprivation	0.01	0.86	0.03	0.86
LSOA Education deprivation	21.97	18.74	22.23	18.86
LSOA Employment deprivation	0.10	0.07	0.10	0.07
Unique GPs	7,635			
Total GP Observations	18,415,832			
Individual Reviews	356,983			

Notes: Summary statistics for data on reviews and enrollment covering GPs in England from May 2015 to December 2021. Total enrollment data is at the GP-quarter level. LSOA enrollment is at the GP-LSOA-Quarter level. An LSOA is a small geographic area with about 700 households (2,000 individuals). The sample excludes GP-LSOA-Quarter observations in the top or bottom 2 percent in terms of quarterly enrollment change.

is 58 patients, and grows by 0.17 patients per quarter on average. The characteristics of registered patients reflect the characteristics of the English population, consistent with the fact that virtually all individuals in England are registered with a GP practice.[16] Panel A of Appendix Figure A-3 shows a histogram of distance between LSOA centroids and chosen GPs. The median individual lives only 1.4 km from their chosen GP, suggesting distance is an important determinant of choice.[17]

[16]The total number of registered patients is actually 4% higher than the population of England. This is thought to be due to over-counting among GPs, under-counting the population, and different definitions of who is a resident. See "Population estimates and GP registers: why the difference?", House of Commons Library, December 12, 2016.

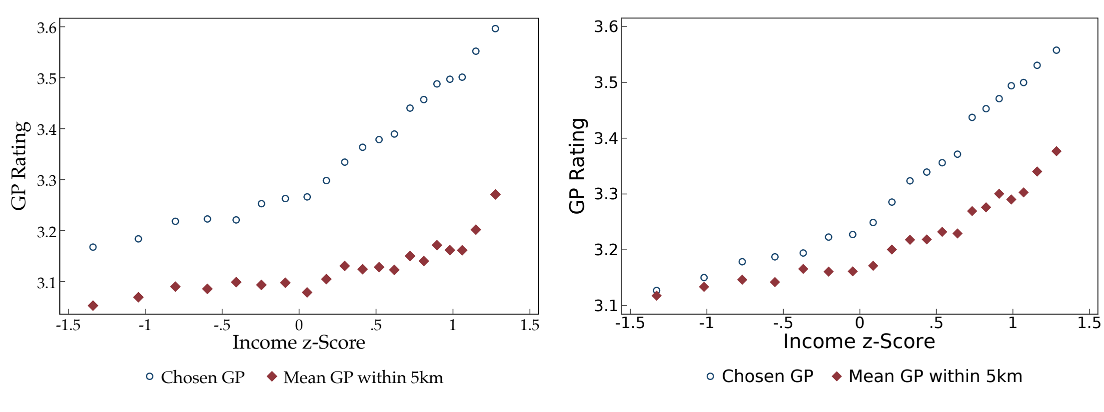
[17]See Panel B of Figure A-3 showing the location of GP practices in England.

2.4 Disparities in GP Quality By Income

We now present descriptive evidence that (i) there are disparities in GP practice quality for those in high versus low income areas and (ii) these disparities can only partially be explained by differences in choice sets. The blue dots in Panel (a) of Figure 1 show the relationship between LSOA level income and the average review index r_{jt}. There is a notable upward trend. Individuals living in LSOAs that are one standard deviation above the mean in terms of the income measure attend GPs that are roughly 0.3 stars higher compared to LSOAs one standard deviation below the mean.

While the star rating system is only one measure of GP quality, there are similar patterns when examining other measures of quality that are arguably more objective. Appendix Figure A-4 shows the relationship between income and the Quality and Outcomes Framework (QOF) clinical score, a measure of GP clinical quality. The relationship between income and this measure of GP quality also shows clear evidence of disparities. This is largely consistent with previous work documenting inequality in health care and a preference for high quality providers in the UK (e.g. Santos et al. 2017; O'Dowd 2020; Scobie and Morris 2020).

Figure 1
Relationship between Income and GP Rating



a. All Patients

b. Movers Only

Notes: Binscatter plots show the relationship between our LSOA-level income measure and the quality of patients' chosen GPs (in blue) as well as the average quality of GPs within 5km (in red). Quality is measured as the average patient rating on the NHS website. Panel (a) includes all patients and Panel (b) includes the sample of patients who moved residences and chose a new GP. The unit of observation is an LSOA-GP-quarter and results are weighted by the quarterly change in enrollment in Panel (a) and enrollment in Panel (b).

The red dots in Panel (a) of Figure 1 suggest that quality disparities cannot be entirely explained by differences in access. These dots show the relationship between LSOA-level income and a proxy for the quality of the choice set faced by individuals in the LSOA: the mean of the average reviews r_{jt} for all GPs within 5 kilometers of the LSOA centroid. Again, there is a noticeable upward trend: high income areas tend to have higher quality GPs nearby. However, the slope is significantly less pronounced than the gradient between income and *chosen* quality. This provides initial evidence that high income individuals choose higher quality GPs even conditional on access. Similar patterns are present when considering various alternative strategies to account for differences in choice sets.

Figure 1 Panel (b) shows that the relationship is even more stark when focusing only on individuals who move residences across LSOAs ("movers"). There is a positive relationship between income (of the destination LSOA) and the quality of patients' newly chosen GP (shown in blue). The upward slope is again much steeper than the corresponding relationship between income and the choice set (shown in red). Notably, movers at the low end of the income distribution choose GP practices that are roughly equivalent to the average within their choice set, while those at the top of the distribution choose GPs that are well above average. In the remaining sections, we explore whether this difference is due to disparities in information, or simply the result of differences in preferences or other factors.

3 A Model of Demand For GP practices with Learning

Our central question is whether differences in information across the income distribution are a meaningful driver of disparities in GP quality. The challenge comes because it is difficult to separate information from preferences or other confounds that lead to heterogeneity in choice by income. In this section, we present a simple model to show that the presence of information gaps—which we define as (differential) imprecision over provider quality—has sharp implications for patient choice in the presence of a star rating website like the one shown on the NHS website. In particular, the model shows that differences in information across groups have distinct empirical predictions from differences in preferences.

The key insight is that patients will only respond sharply to star ratings if they have imprecise private information about quality. Specifically, a Bayesian patient with imprecise private information will have discretely higher beliefs about the quality of practices with better star ratings. On the other hand, the information on the website should have no effect on the choices of an agent that is already perfectly informed. As a consequence, we should see no discontinuous change in demand for these agents across star-rating thresholds, even though they will be responsive to small differences in true quality. This motivates our regression discontinuity analysis in Section 4 and our empirical model in Section 5.

3.1 Patient Beliefs About Practice Quality

Let the quality of GP $j \in \mathcal{J}$ be r_j, and suppose the rounded star rating s_j is public information. We assume that, absent any private information, all agents have prior

$$r_j | s_j \sim \mathcal{N}(\mathbb{E}[r_j | s_j], \eta^2).$$

The prior is centered on the average quality given the observed star rating, with variance η^2 that captures differences in quality across practices that share the same star rating. Now suppose each individual i receives a noisy private signal about the quality of provider j, which could be the result of their own research, recommendations from their social networks, or any other channel. We model this signal as

$$\tilde{r}_{ij} = r_j + \epsilon_{ij}$$

with $\epsilon_{ij} \sim \mathcal{N}(0, \sigma_i^2)$. The signal is centered on the true quality r_j. The variance of the signal (σ_i^2) may differ across individuals due to differential access to information or ability to conduct research.

Given the prior and the signal, individuals form posterior beliefs using Bayes' rule. Expected posterior quality is

$$\mathbb{E}[r_j | \tilde{r}_{ij}, s_j] = \alpha_i \tilde{r}_{ij} + (1 - \alpha_i)\mathbb{E}[r_j | s_j]. \tag{1}$$

The weight individuals place on the private signal is

$$\alpha_i = \frac{\eta^2}{\sigma_i^2 + \eta^2}. \tag{2}$$

Individuals with precise private signals of quality (small σ_i^2) place little weight on the information contained in the star rating, and vice-versa.

3.2 Patient Utility

We assume that patients are risk neutral, value quality r_j, and have a individual taste shocks ν'_{ij}. Given these assumptions, expected utility is

$$\begin{aligned}\mathbb{E}[u_{ij}] &= \beta\mathbb{E}[r_j|\tilde{r}_{ij}, s_j] + \nu'_{ij} \\ &= \beta\alpha_i(r_j + \epsilon_{ij}) + \beta(1-\alpha_i)\mathbb{E}[r_j|s_j] + \nu'_{ij} \\ &= \beta\alpha_i r_j + \beta(1-\alpha_i)\mathbb{E}[r_j|s_j] + \nu_{ij}\end{aligned} \tag{3}$$

where the composite error is $\nu_{ij} = \beta\alpha_i\epsilon_{ij} + \nu'_{ij}$.[18]

3.3 Implications for Ratings and Patient Choice

The model has several predictions about the relationship between GP choice, true quality r_j, and rounded ratings s_j, given a patient's preference for quality (β) and the precision of their information (σ_i). These predictions are summarized in the simulation results shown in Figure 2, which plot demand (choice probabilities) against underlying quality r_j. The simulations are based on a version of the model in which each individual has two providers in her choice set and observes two possible rounded star ratings: $s_j \in \{0, 1\}$.[19] We simulate 1 million individuals, each with a different choice set, and average over the simulations.

Patients with precise private information about quality, which may come because they can easily conduct research on the internet or gather information from their social networks, have a low value of σ_i. For these patients, there is a strong relationship between underlying provider quality r_j and the probability that the provider is chosen. This is visible from the steep slope of the black lines in panel (a) of Figure 2. However, given the precision of their

[18] Since ϵ_{ij} is Gaussian, choice probabilities follow a multinomial probit model if ν'_{ij} is also Gaussian. We assume this in the simulations in Section 3.3.

[19] For these simulations, the taste shock is assumed to be Gaussian. The quality of each GP in an individual's choice set is drawn from a standard normal distribution, with $s_j = 1$ if $r_j > 0$ and 0 otherwise. Note that this specification implies that patients are "behavioral" in the sense of incorrectly believing $r_j|s_j$ is normally distributed. The same predictions and patterns hold if we allow patients to have correct beliefs regarding the truncated distribution $r_j|s_j$, but this complicates the expression in Equation 1 with little additional intuition.

Figure 2
Simulated Demand Response to Star Rating Threshold

a. High signal precision
$\sigma_i = 0.001, \beta = 1$

b. Low signal precision
$\sigma_i = 0.5, \beta = 1$

c. Low preference for quality
$\sigma_i = 0.001, \beta = 0.1$

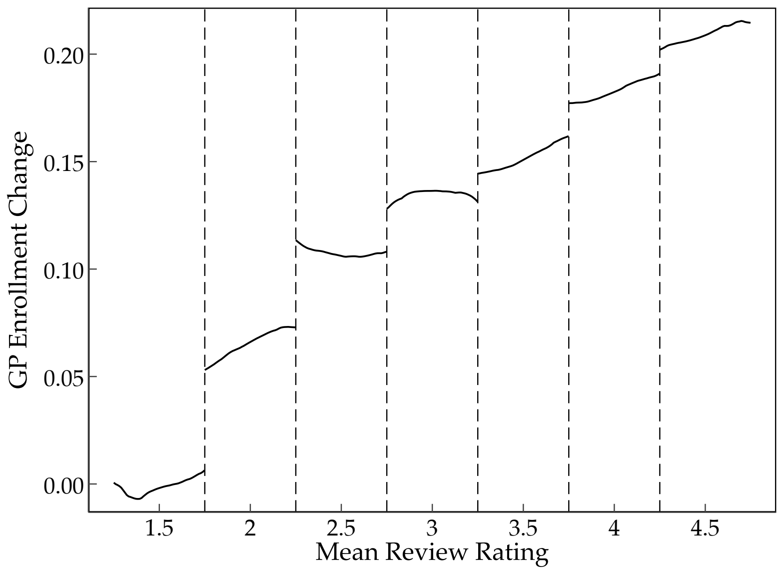
Notes: Simulated demand as a function of quality when individuals receive a signal from star ratings. We assume individuals choose between two GPs with $r_j \sim N(0,1)$. Stars are $s_j \in \{0,1\}$, e.g. can be high or low depending on whether $r_j > 0$. In panel (a), individuals have precise private information about quality and do not respond to the star rating even though they have high demand for quality. In panel (b), individuals do not have precise private signals and have a large response to the star rating. In panel (c), individuals have low demand for quality.

signal, they do not significantly adjust their beliefs on the basis of the star rating itself. This is reflected in the relatively modest jump at the star rating threshold, marked by the black vertical line.

Alternatively, individuals with imprecise private signals (high σ_i) show a relatively flat slope in the relationship between choice probability and r_j away from the threshold, and a sharp discontinuity at the threshold itself. This pattern is shown in panel (b) of Figure 2. Both of these cases are distinct from the results shown in panel (c), which shows a case with precise information, but a low preference for quality (β). In this specification, we observe both a flat slope and a minimal jump at the threshold.

The econometrician can observe both underlying quality r_j and the star ratings s_j even if individuals are not fully informed about r_j. The central takeaway is that the combination of the slope (between demand and r_j) and the jump at the threshold (the response to s_j) allows us to disentangle patients' preferences for quality from the imprecision of their information. This insight guides our reduced form estimation in Section 4. We then expand on and estimate a version of this simple model in Section 5.

Figure 3
GP Enrollment Change and Review Thresholds



Notes: Relationship between average reviews and GP enrollment change by GP-LSOA-quarter for the period when the NHS Choices website displayed star ratings. Lines are smoothed using a local linear regression within each star rating bin. Vertical lines show thresholds for rounding star ratings.

4 RD Evidence on Patient Responses to Star Ratings

Motivated by the model, we examine how low and high income individuals respond to the star ratings provided on the NHS website. We use a regression discontinuity approach that examines jumps in demand at half-star rounding thresholds.

4.1 RD Methodology

Our approach exploits the fact that star ratings s_{jt}, for practice j in quarter t, are rounded to the nearest half star on the NHS Choices website. Two GP practices may display different star ratings even if their average reviews, r_{jt} are very similar. We consider the underlying r_{jt} as our running variable and compare enrollment just above and below each rounding threshold (which exist at every half star interval between 1.25 and 4.75).

Figure 3 displays the basic patterns captured by our RD strategy. We plot GP $\times$ LSOA level changes in enrollment across the distribution of average reviews (allowing separate local linear fits within each value of the star rating s_{jt}). We see a discontinuous upward jump at each rounding threshold, providing initial evidence that patient demand responds to star ratings. Furthermore, we see a consistent, positive, relationship overall, including within each star-rating bin. This indicates that patients prefer GPs with better reviews, even when they share a star rating.

Our main specification collapses the data and jointly estimates the effect of crossing any threshold (because our focus is not on heterogeneity across the range of star ratings). Formally, we let y_{jlt} represent an enrollment outcome for practice j, LSOA ℓ, and quarter t, and let $y_{j\ell t}(s)$ represent the potential outcome with rounded star rating s. Letting c_s be the rounding threshold just above s (i.e. $c_s = s + 0.25$), our target parameter is

$$\tau = \mathbb{E}[y_{j\ell t}(s + 0.5) - y_{j\ell t}(s) | r_{jt} = c_s]. \tag{4}$$

In other words, the average treatment effect of having a one-half-star higher rating, for practices with the index equal to one of the rounding thresholds.

To estimate τ, we stack our data by defining our running variable as $r^0_{jt} = r_{jt} - c_s$ (the distance between the average review and the relevant threshold). This effectively normalizes all thresholds to 0. We take a non-parametric local linear approach following Cattaneo et al. (2019). Specifically, we consider weighted least squares versions of the following regression specification:

$$y_{j\ell t} = \alpha + \tau \mathbb{1}\{r^0_{jt} > 0\} + \beta_- r^0_{jt} + \beta_+ r^0_{jt} \mathbb{1}\{r^0_{jt} > 0\} + X_{jt}\gamma + \varepsilon_{j\ell t}, \tag{5}$$

with weights determined by a kernel function $K(r^0_{jt}/h)$ for a given bandwidth h. Our baseline approach uses a triangular kernel. We include a vector of covariates X_{jt}, which includes GP age, the number of reviews, and cutoff fixed effects. We select symmetric MSE-optimal bandwidths and cluster our standard errors at the GP level using a plug-in residual approach. Unless otherwise noted, we follow Calonico et al. (2019) and the earlier Calonico et al. (2014) when including covariates, selecting bandwidths, and computing standard errors.[20] We include a wide range of robustness exercises to consider alternative bandwidths, alternative kernels, and the exclusion of covariates.

For our RD analysis, our primary outcome variable is the change in enrollment at the

[20]We refer to these bandwidths as CCT bandwidths, given Calonico et al. (2014).

GP-LSOA-Quarter level. To exclude mergers and GP closures, which result in anomalously large jumps in enrollment, we trim observations in which the change in registered patients is in the bottom or top 2 percent. Our results are robust to alternative methods for identifying mergers and closures. We also exclude practices with fewer than 5 reviews in our primary RD analysis, because star ratings for GPs with very few reviews carry limited information.[21] Furthermore, we drop practices with an average review exactly equal to one of the rounding thresholds. We include robustness exercises that relax these restrictions.

Our analysis proceeds in three steps. First, we discuss the plausibility of our identifying assumptions. Second, we show that rounded star ratings impact enrollment, on average, in the population. Finally, we show that these impacts are primarily driven by patients living in low income LSOAs.

4.2 Plausibility of Identifying Assumptions

Our key identification assumption is that the relevant average potential outcomes functions are continuous at each threshold.[22] This may fail if there is endogenous sorting of GPs at rounding thresholds, or if the observable or unobservable characteristics of GPs change sharply for other reasons. For instance, a number of GPs might exert effort to stay just above a threshold. Alternatively, some more fundamental features of the rating system (e.g, the number of reviews given to particular types of providers or a tendency of certain GPs to manipulate reviews), could generate discontinuities. Below, we show three pieces of evidence in support of our continuity assumption.

Manipulation Tests. We perform manipulation tests in the spirit of McCrary (2008) to rule out jumps in the distribution of the running variable r_{jt} across the threshold. We implement the tests outlined in Cattaneo et al. (2018) based on local polynomial density estimators. We use the suggested MSE optimal bandwidth and show unrestricted robust bias-corrected t-statistics. Using the analysis sample described above, we find no evidence of a discontinuity in the density of average reviews at the threshold ($t = -1.15$). We present our results in Appendix Figure B-1. This suggests that practices in our sample did not differentially

[21] Average ratings are less correlated with other measures of quality when the number of reviews is small (see Appendix Table A-1), and individuals are likely aware of this.

[22] That is, $\mathbb{E}[y_{j\ell t}(s)|r_{jt}]$ and $\mathbb{E}[y_{j\ell t}(s+0.5)|r_{jt}]$ are continuous at $r_{jt} = c_s$ for each star rating s.

manipulate reviews (either falsely or through the provision of effort) to gain a higher star rating, and rules out other sources of bunching in the distribution of practices.

Smoothness of Covariates. We test the smoothness of covariates across the thresholds, to rule out discrete jumps in the observable features of GPs above versus below each threshold. Appendix Figure B-2 shows that observable characteristics of GPs are continuous at rounding thresholds. For our tests, we implement the RD methodology described in Equation 5 above at the GP level with five different GP characteristics on the left-hand side: (i) the fraction of patients that report having to wait one week or more for an appointment (in survey data) (ii) the number of months the GP has been active, (iii) a survey based outcome of patient trust, (iv) the QOF clinical score, and (v) the payments each GP receives per-patient from the NHS. In each panel, we report the t-test for the null that $\tau = 0$. We also show binned scatter-plots representing the means of each variable above and below the threshold, as well as estimates from local linear regressions. The average values of all are smooth through the threshold, with small t-statistics, suggesting no discrete change in provider type. Note in particular there is no detectable change in patient wait times at the threshold, indicating that differential overcrowding is not the source of our result.

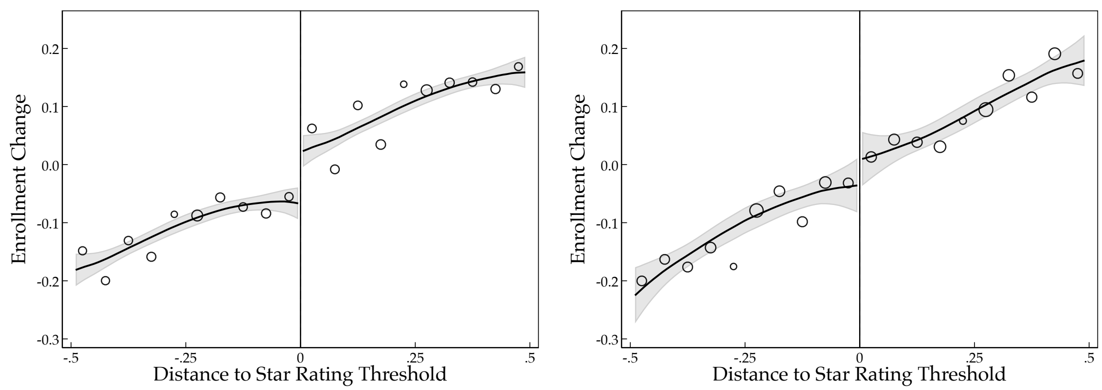
Distribution of Patient Ratings. To further test for manipulation of the star rating system, we take advantage of the period after 2020, in which patients could still leave reviews but summary star ratings were not displayed. If providers were actively gaming star ratings, we would expect a discrete change in the distribution of reviews after the stars were taken offline (and the incentive to manipulate disappeared). Appendix Figure B-3 shows that there was no statistically significant difference in the distribution of reviews before versus after the star ratings were displayed, providing evidence against fake reviews or strategic behavior on the part of GPs.

4.3 RD Results

Average Impact of Star Ratings on Enrollment

The baseline impact of star ratings on enrollment using our full sample is shown graphically in Panel (a) of Figure 4. The chart shows enrollment change as a function of the distance to

Figure 4
Effect of Star Rating Threshold on GP Enrollment
Before and After Website Change



a. Visible star ratings

b. No star ratings

Notes: Mean enrollment change around the threshold for rounding star ratings. The size of the circles corresponds to the number of observations in each bin. The fitted line is from a local linear regression using a triangular kernel. Shaded area shows 95% confidence interval.

the star rating threshold. We show both a binned scatter-plot and a local linear smoothing on each side of the threshold. The relationship between average reviews and the change in enrollment is roughly linear with a positive slope on both sides. This slope indicates that on average patients value—and have some private information about—the notion of quality captured by the running variable r_{jt}. There is a sharp jump in enrollment at the star rating threshold, indicating that a higher star rating has a meaningful effect on demand.

The RD results corresponding to this figure are presented in the first two columns of the top panel of Table 2. Our preferred estimates (column 1) imply that a half star jump in ratings increases the quarterly change in enrollment from an LSOA by 0.13 patients. The mean LSOA quarterly enrollment change at a GP practice is 0.17, implying that an additional half star increases quarterly enrollment by 75% of the mean. This effect is statistically significant at the 5% level. Column 2 shows that results are robust to using the bandwidth selection procedure of Imbens and Kalyanaraman (2012). Our results suggest that patients value and respond to the information provided by the star-rating system.

As a falsification test, we examine the period from January 2020 to December 2021, when rounded star ratings were no longer shown on the NHS website. If stars are responsible for the jumps shown in Panel (a) of Figure 4, we should expect the jumps to disappear after

Table 2
Effect of Star Ratings on Enrollment Change
Regression Discontinuity Estimates

	Visible Star Ratings		No Star Ratings	
	CCT Bandwidth	IK Bandwidth	CCT Bandwidth	IK Bandwidth
Estimate	0.131	0.073	0.030	0.031
	(0.058)	(0.034)	(0.105)	(0.061)
P-Value	0.025	0.031	0.775	0.606
Robust CI	[.009 ; .278]	[.019 ; .206]	[-.228 ; .282]	[-.148 ; .24]
Bandwidth	0.13	0.39	0.13	0.30
N	916,822	2,801,989	310,307	716,328
	Visible Star Ratings		No Star Ratings	
	Low Income	High Income	Low Income	High Income
Estimate	0.185	0.058	-0.098	0.153
	(0.068)	(0.072)	(0.140)	(0.139)
P-Value	0.007	0.424	0.482	0.271
Robust CI	[.05 ; .359]	[-.1 ; .238]	[-.479 ; .179]	[-.133 ; .524]
Bandwidth	0.15	0.12	0.11	0.12
N	507,107	427,664	138,215	140,707
T-Test by Income		2.64		-1.44

Notes: Dependent variable is the change in enrollment at the GP-LSOA-Quarter level. Sample criteria described in Subsection 4.1. Visible star ratings indicates the period prior to January 2020. No star ratings indicates January 2020 and later. Low (high) income is defined as LSOAs below (above) the median of our income measure. RD estimates use local linear regressions with triangular kernels. controls for GP age, age squared, and number of practitioners in the GP practice, as well as threshold fixed effects are included. We follow Calonico et al. (2019) and Calonico et al. (2014) to select bandwidths, include covariates, calculate standard errors clustered at the GP level (in parentheses), and construct robust bias corrected confidence intervals. For columns labeled IK, bandwidths are selected following Imbens and Kalyanaraman (2012).

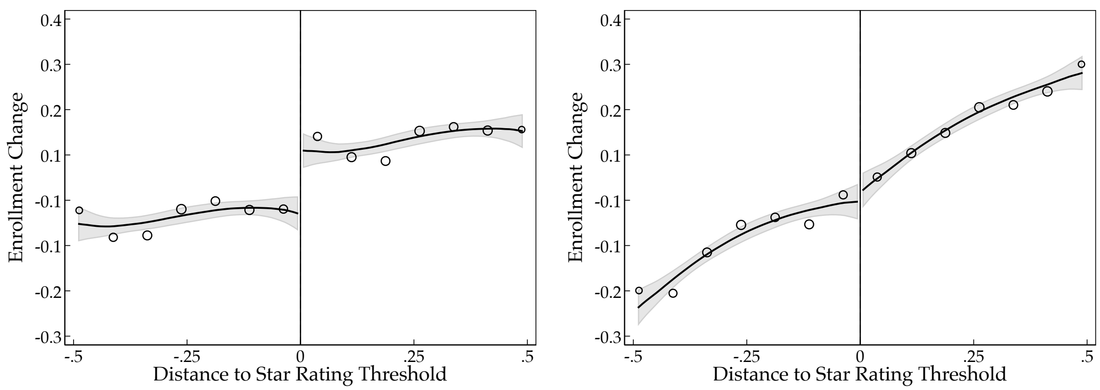
the ratings were removed. Indeed, Panel (b) of Figure 4 shows that there is no effect at the threshold during this period even though the relationship between reviews and GP demand is similar on either side. Our RD estimates corroborate this pattern. There is no statistically significant effect in this period (columns 3 and 4 of Table 2).[23]

[23]We similarly find no effect when examining only the first quarter of 2020, prior to the widespread COVID-19 pandemic, which suggests this null effect is not a spurious consequence of the pandemic.

Differences by Income

We now turn to heterogeneity in the impact of star ratings by income. We separately implement our RD strategy on subsamples of high and low income LSOAs. High and low income are defined as those above versus below the median of our LSOA level income measure.

Figure 5
Effect of Star Rating Threshold on GP Enrollment by Income



a. Low Income　　　　b. High Income

Notes: Mean enrollment change around threshold for star ratings, splitting at median LSOA income, in the period in which star ratings were visible. The size of the circles corresponds to the number of observations in each bin. The fitted line is from a local linear regression using a triangular kernel. Shaded area shows 95% confidence interval.

Figure 5 presents graphical evidence of our main result: the impacts of the star rating website are largely driven by patients in low income LSOAs. Panel (a) shows evidence of a sharp discontinuity in demand for LSOAs with below-median income. There is no evidence of a discontinuity in Panel (b), which shows LSOAs with above-median income. The regression estimates in the first two columns of panel (b) of Table 2 confirm this result. For individuals in LSOAs with below-median income, the effect of crossing a star ratings threshold on enrollment change is 0.19 and is significant at the 1% level. The point estimate for LSOAs with above-median income is much smaller (0.06) and is not statistically significant. Furthermore, the discontinuities for high and low income LSOAs are statistically different from one another (t-test statistic=2.64). As the last two columns of Panel (b) show, neither result is significant in the period without star ratings, providing additional validation for our result. Together, this evidence suggests that demand for low income patients is more responsive to the information contained in star ratings.

While not the focus of our RD approach, the relationship between demand and average ratings away from the threshold provides further insight into preferences for quality, as highlighted by the model in Section 3. For high income LSOAs, we see relatively steep slopes on either side of the threshold, despite the lack of a discontinuity at the threshold itself. This suggests that high income individuals value the measure of GP quality captured by r_{jt} and have significant information on r_{jt} beyond what is provided by the website. In contrast, we see relatively flat slopes away from the threshold for low income LSOAs, despite the large jump in demand at the threshold. This implies that low income individuals also value GP quality, but rely heavily on the website for their information.

Through the lens of our model, these results suggest that both low and high income patients value GP quality, but that high income patients have significantly more precise information. In other words, that there are information gaps by income. There are a series of potential alternative interpretations for these patterns, which we now discuss in turn.

Alternative Interpretations

A first potential issue is that low income patients simply change providers more frequently, and the sizeable RD coefficients we estimate are an artifact of this difference in levels. This does not appear to be the case. Average enrollment changes are similar across income groups. Furthermore, in Appendix Tables B-1 and B-2, we repeat our analysis focusing on enrollment growth rates, which account for baseline differences in levels. We again find significantly larger jumps for low income LSOAs. The patterns are similar when using individual-level data and limiting the sample to "movers," individuals who relocated their primary residence to a different LSOA and chose a new GP. The results, displayed in Appendix Figure B-4 and Panel B of Appendix Table B-2, also show a differential response by income, implying that differential inertia is not driving our findings.

A second concern is that low income patients care more about the measure of quality captured by r_{jt}, and hence are more responsive to star ratings. This interpretation is difficult to square with the steep upward slope away from rounding thresholds that we find for high income patients. Overall, the choices of high income individuals are highly correlated with the r_{jt} despite the fact that they do not respond at the threshold. In fact, as Figure 5 shows, the total change in demand from a half star below to a half star above the threshold is more than twice as large for high income LSOAs (versus low income LSOAs). We explicitly disentangle preferences for quality from information in our structural estimation, but the basic patterns suggest that high income patients care about, and respond to, this measure of quality.

A third concern is that our RD estimates reflect a greater skill in navigating technology and the star rating website among low income patients. This explanation is at odds with standard views of the correlation between income and technological skill, which typically suggest the opposite.[24] A related possibility is that our measure of income is just a proxy for age, and our RD estimates reflect the relative technological sophistication of younger patients. We rule this out in Appendix Table B-3, which shows that low income patients are more responsive at rounding thresholds even *within* age categories.

A fourth concern is that some feature of the set of choices faced by low income patients generates a particularly large response at the threshold. This might occur if star-rating treatment effects are larger at lower thresholds (and low income patients are more likely to face these thresholds), or if capacity constraints bind for practices just above thresholds in high income neighborhoods, discouraging new patients from enrolling in large numbers. To address this possibility, the first two columns of Appendix Table B-4 consider only high and low income patients living in close proximity, who therefore face similar choice sets. RD estimates are similar to our baseline specifications. To examine the role of capacity constraints, the remaining two columns of Appendix Table B-4 show a specification that excludes practices that appear to be constrained.[25] We again find similar results.

4.4 Robustness of RD Approach

Appendix Tables B-1 and B-2 show that our RD approach is robust to our choices of bandwidth and kernel, and that our results are not driven by our sample selection choices, the inclusion of covariates, or the particular measure of enrollment used. Across all specifications, we estimate discontinuous jumps in enrollment at rounding thresholds and find that

[24]For example, see the discussion of technology in the UK Social Mobility Commission's 2021 *State of the Nation* report (Commission 2021).

[25]Our proxy for constraints is based on consecutive months with similar total enrollment.

these jumps are driven by residents of low income LSOAs.

We also find similar results using a complementary panel fixed-effects strategy that exploits within-GP changes in the rounded star rating s_{jt} over time. These changes occur as new reviews are added, or as older reviews fall out of the two-year moving average. We implement this approach by regressing enrollment changes on the star rating and the star rating interacted with LSOA income, while controlling for GP fixed effects, quarter×year fixed effects, and time-varying GP controls. While this strategy relies on stronger identification assumptions than our RD approach, we show that low income patients are more responsive to *changes* in star ratings. We present the estimation details and results in Appendix Section C.

Because inequality by *income* in health care (and other outcomes) is salient and widely discussed, we use income as our key dimension of heterogeneity. However, we do not intend to suggest that the role of income is causal, and we recognize that it is likely a proxy for some more complex underlying factor. In Appendix Figure B-5 we show evidence that similar patterns hold across other measures of socioeconomic status that are highly correlated with income. Individuals with below-median education, employment, or health all respond to the star ratings. Individuals with above median education, employment, or health do not have a statistically significant response at the threshold, but do demand quality (have steep slopes) away from the threshold. We also show more fine-grained evidence on the income distribution in Appendix Figure B-6, which shows separate plots by income quartile: those in the lowest quartile respond most at the star rating thresholds.

5 Empirical Model of Demand

We now present an empirical framework that extends the simple model in Section 3.1. The model allows us to quantify the role of information, preferences, and access as drivers of disparities by explicitly incorporating variation in information generated by rounded star ratings (following past literature that has leveraged variation in information, e.g., Brown 2019; Allende et al. 2019; Brown and Jeon 2022).

5.1 Empirical Model Setup

Beliefs

At time t, all individuals share a prior belief about r_{jt}, the quality of GP j, conditional on the website's star rating s_{jt}, which is given by

$$r_{jt}|s_{jt} \sim \mathcal{N}(m_{jt}, \eta^2).$$

Where the mean and variance of beliefs are based on the empirical distribution of reviews.

Individual i in LSOA ℓ and quarter t receives a noisy private signal about the quality of GP j given by $\tilde{r}_{i\ell jt} = r_{jt} + \epsilon_{i\ell jt}$, where

$$\epsilon_{i\ell jt} \sim \mathcal{N}(0, \sigma_\ell^2).$$

To capture differences in information, we allow the precision of private information to vary across LSOAs. Specifically, we parameterize the precision of each individual's private signal as a function of LSOA-level income:

$$\frac{1}{\sigma_\ell^2} = \exp[\gamma_0 + \gamma_1 I_\ell].$$

The mean of posterior beliefs about the quality of GP j for individual i in LSOA ℓ and quarter t is therefore

$$\mathbb{E}[r_{jt}|\tilde{r}_{i\ell jt}, s_{jt}] = \alpha_\ell(r_{jt} + \epsilon_{i\ell jt}) + (1 - \alpha_\ell)m_{jt} \tag{6}$$

where the weight that patients put on their private signal, which also varies across LSOAs, is

$$\alpha_\ell = \frac{\eta^2}{\sigma_\ell^2 + \eta^2}. \tag{7}$$

In other words, patients put more weight on the star ratings when their private signals are noisy.

Utility

Expected utility from choosing GP $j \in \mathcal{J}_{\ell t}$ is given by

$$\begin{aligned}\mathbb{E}[u_{i\ell jt}] &= \beta_{1\ell}\mathbb{E}[r_{jt}|\tilde{r}_{i\ell jt}, s_{jt}] + f(d_{\ell j}; \beta_{2\ell}) + X'_{jt}\beta_3 + \xi_j + \nu'_{i\ell jt} \\ &= \beta_{1\ell}[\alpha_\ell r_{jt} + (1-\alpha_\ell)m_{jt}] + f(d_{\ell j}; \beta_{2\ell}) + X'_{jt}\beta_3 + \xi_j + \nu_{i\ell jt}. \end{aligned} \tag{8}$$

Representative utility is a function of the mean of posterior beliefs about quality $\mathbb{E}[r_{jt}|\tilde{r}_{i\ell jt}, s_{jt}]$, given by equation 6. We let preferences for quality vary with income by assuming $\beta_{1\ell} = \beta_{1,0} + \beta_{1,1} I_\ell$. Utility is also a function of the distance $d_{\ell j}$ between the individual's LSOA ℓ and GP j. We parameterize the disutility of distance as $f(d_{\ell j}; \beta_{2\ell}) = \beta_{2\ell}\log(d_{\ell j})$ and assume $\beta_{2\ell} = \beta_{2,0} + \beta_{2,1} I_\ell$, so we allow the disutility of distance to vary by LSOA income.

Utility additionally depends on other GP characteristics X_{jt}, which includes mean physician age within the practice and the number of practitioners per patient as a proxy for wait times or capacity constraints. ξ_j are GP fixed effects capturing other dimensions of time-invariant quality not captured by the reviews or observable characteristics (e.g., convenient location).

Inertia

A modest fraction of individuals switch providers in any given quarter, in line with many other healthcare choice settings featuring inertia (e.g., Raval and Rosenbaum 2018; Shepard 2022). We model this as an exogenous re-optimization probability φ_ℓ. This can be thought of as jointly capturing any factors that cause individuals to consider selecting a new GP, such as health shocks or moving.

We parameterize inertia as

$$\varphi_\ell = \frac{1}{1 + \exp[-(\theta_1 + \theta_2 I_\ell)]}$$

so that the probability that an individual re-optimizes may vary by LSOA income.

Market Shares

Following our theoretical model, the composite error $\nu_{i\ell jt} = \beta_{1\ell}\alpha_\ell\varepsilon_{i\ell jt} + \nu'_{i\ell jt}$ has two components. The first ($\beta_{1\ell}\alpha_\ell\varepsilon_{i\ell jt}$) is driven by the noise in individual i's private signal about quality of GP j. The second ($\nu'_{i\ell jt}$) represents individual i's taste shock and is normalized to have unit variance, so $\mathbb{V}[\nu_{i\ell jt}] = \beta_{1\ell}^2\alpha_\ell^2\sigma_\ell^2 + 1$. For computational tractability, we assume that the composite error $\nu_{i\ell jt}$ follows an iid EV1 distribution. It is useful to define

$$k_\ell^2 = \frac{6}{\pi^2}(\beta_{1\ell}^2\alpha_\ell^2\sigma_\ell^2 + 1) \tag{9}$$

such that $\nu_{i\ell jt}/k_\ell$ is distributed iid EV1 with scale parameter 1 (and therefore has the usual variance $\pi^2/6$). Conditional on making an active choice, choice probabilities are then

$$p_{\ell jt} = \frac{\exp\left[\frac{1}{k_\ell}\left(\beta_{1\ell}\left[\alpha_\ell r_{jt} + (1-\alpha_\ell)\mathbb{E}[r_{jt}|s_{jt}]\right] + f(d_{\ell j};\beta_{2\ell}) + X'_{jt}\beta_3 + \xi_j\right)\right]}{\sum_{k\in\mathcal{J}_{\ell t}}\exp\left[\frac{1}{k_\ell}\left(\beta_{1\ell}\left[\alpha_\ell r_{kt} + (1-\alpha_\ell)\mathbb{E}[r_{kt}|s_{kt}]\right] + f(d_{\ell k};\beta_{2\ell}) + X'_{kt}\beta_3 + \xi_k\right)\right]}. \tag{10}$$

The variance of the composite error ($\nu_{i\ell jt}$) differs across LSOAs due to heterogeneity in the precision of the private signal (σ_ℓ^2). The term $1/k_\ell$ normalizes the scale of the error so that this variance is constant. Notice that k_ℓ depends on $(\beta_{1\ell}, \sigma_\ell)$ and thus will be estimated.

Given this, and the assumptions on individual inertia described above, the market share predicted by the model for GP j within LSOA ℓ in quarter t is

$$S_{\ell jt} = \varphi_\ell p_{\ell jt} + (1-\varphi_\ell)S_{\ell j,t-1} \tag{11}$$

where $S_{\ell j,t-1}$ is the market share of practice j in period $t-1$. With probability $1-\varphi_\ell$ patients do not re-optimize, and instead remain in the same GP they were enrolled within period $t-1$. Otherwise, patients re-optimize and choose according to the choice probabilities $p_{\ell jt}$.

5.2 Estimation Approach

We estimate four types of parameters, which represent preferences for observables, GP fixed effects, consumer information, and inertia, ($\Phi = \{\beta, \xi, \gamma, \theta\}$). We use an indirect inference approach in which parameters of the model are chosen such that results from an auxiliary model using data simulated from our demand model are as close as possible to results from the same auxiliary model using the observed data Gourieroux et al. (1993). We specify our auxiliary model based on the regression discontinuity strategy outlined in Section 4.1. Our approach is related to Fu and Gregory (2019) and Allende (2022) who integrate RD estimates into empirical models of post-disaster subsidies and school choice, respectively. As described below, we also target additional moments based on shares and switching rates.

Moment Conditions

Given a guess of parameters Φ, the model predicts market shares $\widetilde{S}_{\ell jt}(\Phi)$. We use these market shares to construct four sets of moment conditions described below.

The first set of moments matches the RD estimates using simulated data implied by the model to the RD estimates obtained from the observed data. Using the model predicted enrollment changes for each GP×LSOA pair, we estimate a regression discontinuity following Equation 5, as described in Section 4.1. For computational tractability, we specify a uniform kernel and estimate using a standard OLS regression. For simplicity, we allow for a single slope parameter above and below the threshold (i.e. fix the parameter $\beta_+ = 0$).[26] This provides estimates of the jump at the threshold, as well as the slope on either side of the threshold.

Let $I^{(2)} \in \{1, 2\}$ index LSOA income (i.e., above and below median). We estimate these RD regression coefficients separately for each income level $I^{(2)}$ to mirror the analysis of Section 4.3. When computing these RD estimates, we also follow the sample restrictions used for the results in Section 4.3 (for all other moments, we use the full estimation sample described below). Defining the vector of slope and jump coefficients implied by the simulated data as $\widetilde{\tau}_{I^{(2)}}(\Phi)$, we consider the moments

$$M^1(\Phi) = \widetilde{\tau}_{I^{(2)}}(\Phi) - \widehat{\tau}_{I^{(2)}} \quad \forall \ I^{(2)}. \tag{12}$$

Here, $\widehat{\tau}_{I^{(2)}}$ represents the RD slope and jump estimates using the observed data.

The second set of moments matches model implied market shares to observed market shares. The error in LSOA-year-GP-level market shares is $\widetilde{S}_{\ell jt}(\Phi) - S_{\ell jt}$, where $S_{\ell jt}$ are market

[26] In practice, estimates from this model and our more flexible non-parametric RD model are similar. See Appendix Table B-5.

shares in the data. There are roughly 2 million GP-quarter-LSOA observations in the sample used to estimate the model. For computational tractability, we average across LSOAs at the GP-year level, implying 7,222 moments. Let $\mathcal{G}_{jt}$ be the set of LSOAs where GP j is in the choice set for quarter t and define L_{jy} as the number of LSOA-quarter pairs where GP j is a member of choice set during calendar year y. We consider

$$M^2(\Phi) = \frac{1}{L_{jy}} \sum_{t:\, y(t)=y} \sum_{\ell \in \mathcal{G}_{jt}} (\widetilde{S}_{\ell jt}(\Phi) - S_{\ell jt}) \quad \forall\, j, y. \tag{13}$$

The third set of moments matches the average value of observable characteristics $\mathbf{x}_{\ell jt}$ of chosen GPs to the averages implied by predicted market shares, as in the "micro moments" commonly used in demand estimation (Petrin 2002; Berry et al. 2004):

$$M^3(\Phi) = \sum_{\ell,j,t} (\widetilde{S}_{\ell jt}(\Phi) - S_{\ell jt}) w_{\ell t} \mathbf{x}_{\ell jt}. \tag{14}$$

The vector $\mathbf{x}_{\ell jt}$ represents all observable variables that enter utility. This includes log distance, distance interacted with LSOA income, each component of posterior beliefs about quality (r_{jt}, m_{jt}, and each interacted with income), GP experience, and practitioners per patient. The weight $w_{\ell t}$ is the share of total enrollment in the full sample accounted for by the LSOA-quarter.

The fourth set of moments attempts to match the model predictions for how often individuals switch GPs to the observed switching rates in individual-level data. We match these switching rates separately by LSOA income quartile. Let $I^{(4)} \in \{1, 2, 3, 4\}$ index quartiles of LSOA income. Given the observed fraction of patients that change practice each quarter on average, for the LSOAs in each quartile of income (which we label $switchrate_{I^{(4)}}$), model predicted choices probabilities $\widetilde{p}_{\ell jt}(\Phi)$, observed lagged enrollment $S_{\ell j,t-1}$, and lagged market size $T_{\ell,t-1}$, the difference between the aggregate switching rate implied by the model and the observed data is:

$$M^4(\Phi) = \frac{\sum_{\ell jt} \varphi_{\ell t}(1 - \widetilde{p}_{\ell jt}(\Phi)) S_{\ell j,t-1} T_{\ell,t-1} \mathbb{1}(I_\ell^{(4)} = I^{(4)})}{\sum_{\ell jt} S_{\ell j,t-1} T_{\ell,t-1} \mathbb{1}(I_\ell^{(4)} = I^{(4)})} - switchrate_{I^{(4)}} \quad \forall\, I^{(4)} \tag{15}$$

In the model, switching is the joint probability of (i) making an active choice, which occurs with probability $\varphi_{\ell t}$ and (ii) selecting a practice *other than* the current practice j, which occurs with probability $1 - \widetilde{p}_{\ell jt}(\Phi)$.

We then stack the moment conditions $M^1(\Phi), M^2(\Phi), M^3(\Phi), M^4(\Phi)$ into a vector of moments M. Our estimator is the solution

$$\widehat{\Phi} = \text{argmin}_{\Phi}\, M'WM$$

where W is a positive definite weighting matrix.

Estimation Sample

Given computational constraints, our estimation focuses on the sample of LSOAs in greater London.[27] The RD results for this sample are similar to those for all of England.[28] We define the choice set for each LSOA-quarter ($\mathcal{J}_{\ell t}$) as the set of active GPs within 3km of the LSOA centroid. This includes some observations with zero enrollment, although we exclude GP-LSOA pairs with zero enrollment over the entire sample as well as practices that have closed and are inactive in all LSOAs as of quarter t.

Several elements of the model are measured directly from the data. We measure r_{jt} using the two-year rolling average of reviews, which is what determines the star ratings. Lagged market shares $S_{\ell j,t-1}$ are directly observed from data. The information contained in star ratings is inversely proportional to η^2, which is calculated directly from the data as the variance of quality r_{jt} within each star rating bin, averaged across all bins.

To account for the fact that demand is not very responsive to star ratings when the number of reviews is very small, we apply a shrinkage procedure to the mean of prior beliefs, m_{jt}. In short, for a practice with few reviews, m_{jt} is shrunk to the average quality across all practices.[29] This primarily affects expected quality when the number of reviews is

[27] Computational time scales rapidly with the number of GP practice fixed effects ξ_j. By restricting the sample geographically, we are able to reduce the number of GPs being considered.

[28] See Table 2 and Appendix Table B-5. As in our baseline results, low income individuals respond more at the star rating thresholds than high income individuals, and demand by high income individuals is more responsive to the underlying average reviews within each star-rating bin. In Greater London, there is also a similar relationship between average quality of chosen GPs and income (Appendix Figure A-5).

[29] Let n_{jt} be the number of reviews, $\mathbb{E}[r_{jt}|s]$ be the mean of reviews conditional on a rating s across all periods, and $\mathbb{E}[r_{jt}]$ be the mean of reviews across all practices and periods. Let ψ^2 be the variance of *individual reviews* (i.e., postings to the website with values in $\{1, 2, 3, 4, 5\}$) for all practices over all periods. Let ψ_s^2 be the variance of individuals' reviews across all practices and periods conditional on a rating s. We assume the mean of prior

very low, i.e., less than 5.

Identification

A key challenge is separate identification of the preference for quality ($\beta_{1\ell}$), the precision of individuals' private information (σ_ℓ), and GP unobserved quality (ξ_j). In general, it is difficult to determine if individuals do not choose high quality GPs because they do not know they are high quality or they do not have a strong preference for quality. However, as discussed in Section 3.3, information and preferences have different implications for how individuals respond to changes in star ratings in our setting.

First, consider a GP for which r_{jt} increases over time but does not cross a threshold. Since the GP remains within the same star-rating bin, the public information regarding this GP does not change. Given that ξ_j is constant, a higher correlation between r_{jt} and demand implies a higher value of $\beta_{1\ell}$. This argument uses only time-series variation within each GP. In practice, conditional on $\beta_{1\ell}$ and ξ_j, cross-sectional correlation between demand and r_{jt} across providers within a star-ratings bin also informs the estimation of $\beta_{1\ell}$.

Second, consider a GP for which r_{jt} increases over time in such a way that the star rating changes. If the increase in demand is larger than what would be expected given a fully informed individual, this implies that individuals have imprecise private information (σ_ℓ is large). To the extent that low income individuals respond more when a GP crosses a threshold, this helps pin down how the precision of private signals varies with income. This is similar to the identification in the panel regression approach in Appendix Section C. Again, the cross-sectional variation used in the RD analysis also informs the estimation of σ_ℓ. Conditional on $\beta_{1\ell}$ and on the distribution of ξ_j being smooth across thresholds, the magnitude of the RD estimates provide information about σ_ℓ.[30]

How each moment informs estimation can be summarized as follows. Moments $M^1(\Phi)$ target the RD estimates, therefore informing the parameters governing the precision of information (σ_ℓ) as well as the preferences for quality ($\beta_{1\ell}$). Moments $M^2(\Phi)$ target market shares averaged across LSOAs at the GP-quarter level. These moments help inform the estimation

beliefs is $m_{jt} = \omega_{jt}\mathbb{E}[r_{jt}|s_{jt}] + (1-\omega_{jt})\mathbb{E}[r_{jt}]$ where $\omega_{jt} = \psi^2/(\psi^2 + \frac{1}{n_{jt}}\psi_s^2)$. When constructing counterfactuals, we fix $\omega_{jt} = 1$.

[30] After estimation, we verify that ξ_j are smooth across the RD thresholds. See Appendix Figure D-1.

of the GP fixed effects ξ_j. Intuitively, if GP j tends to have large market shares across multiple LSOAs and time periods, $M^2(\Phi)$ is minimized by a large value of ξ_j. Moments $M^3(\Phi)$ target market shares weighted by provider characteristics. These moments primarily inform the coefficients on distance ($\beta_{2\ell}$) and other provider characteristics (β_3). This also helps ensure that the model can capture the fact that preferences for characteristics like distance may vary by income. Moments $M^4(\Phi)$ target the switching rates in the data, therefore primarily informing the estimation of the inertia parameters and heterogeneity in inertia by income (θ_1, θ_2).

5.3 Model Estimates

We present the estimates from our structural model in Table 3. Appendix Section E describes how we compute standard errors for our GMM estimator. We focus on two sets of parameters. First, γ_0 and γ_1, which govern the precision of the private signal. Our model estimates are in line with the implications of our reduced form evidence. On average, patients have noisy private information about GP quality, but higher income patients have significantly more precise information. The estimates imply that the weight placed by individuals on their private signal (α_ℓ) has an average of 0.57. However, the role of income heterogeneity is significant. Individuals from LSOA's in the 10th percentile of income have an $\alpha_\ell \leq 0.06$, suggesting they put nearly all weight on the website's star ratings. At the 90th percentile of income, $\alpha_\ell \geq 0.95$, indicating that almost all weight is placed on their private signal.

The second key set of parameters, $\beta_{1,0}, \beta_{1,1}$, govern preferences for GP quality. Both high and low income patients have a large and significant preference for higher quality GPs. Higher income individuals have a slightly stronger preference for quality, but the coefficient on income is not statistically significant. The parameters governing preferences for other observable GP characteristics are reasonable. For example, patients dislike GPs that are far from their LSOA, and this distaste is somewhat weaker for high income individuals. Given that high income individuals may be willing to travel further to a high quality GP, this may partially explain disparities in quality. As expected, individuals prefer GP practices with a higher number of practitioners per patient, suggesting a dislike for congestion or longer waiting times.

Table 3
Estimates for GP Demand Model

	Estimate	SE
Inertia (θ)		
Constant	−3.406	(0.002)
Income	0.095	(0.002)
Private Signal Precision ($\frac{1}{\sigma_\ell^2}$)		
Constant	4.313	(0.572)
Income	2.214	(0.617)
GP Quality ($\beta_{1\ell}$)		
Constant	0.284	(0.020)
Income	0.011	(0.021)
Distance ($\beta_{2\ell}$)		
Constant	−1.778	(0.028)
Income	0.036	(0.029)
Other GP Characteristics (β_3)		
Mean physician age	0.049	(0.026)
Practitioners per 1000 Patients	0.224	(0.046)
Active choice fraction ($\bar{\varphi}_\ell$)	0.032	
Private Signal Weight (α_ℓ)		
Mean	0.565	
10th	0.057	
25th	0.227	
50th	0.636	
75th	0.886	
90th	0.957	

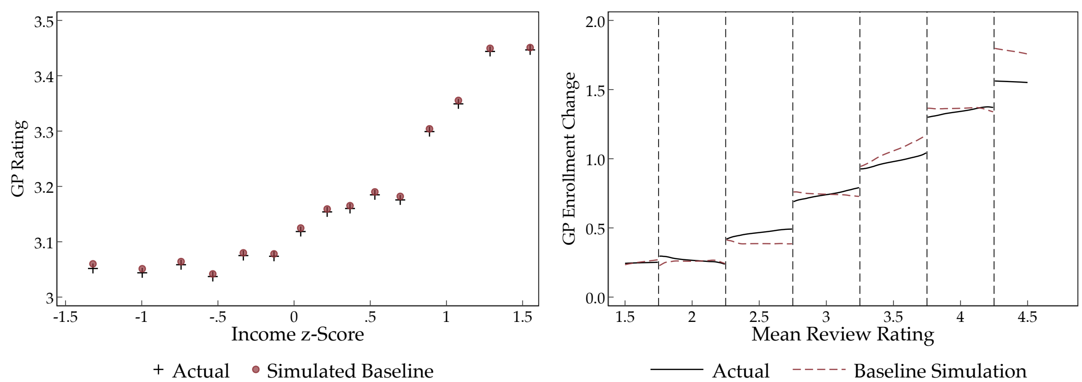
Notes: Top panel presents parameters from demand model estimated via indirect inference. Appendix Section E provides details on how standard errors are computed. Bottom panel shows average fraction of individuals making an active choice and distribution of signal weights at the estimated parameter values.

5.4 Model Fit

We are able to closely match our targeted moments. The RD estimates implied by our auxiliary model match the RD estimates implied by the data to three decimal places. We also match market shares well, with a mean absolute percentage error of 3%. The model also matches observed switching rates, including the fact that high income individuals switch at slightly higher rates. Overall, the model implies that 3.2% of patients make an active choice in any given quarter. We are similarly effective in matching observable characteristics. Appendix Figure D-2 and Table D-1 summarize our model fit with respect to targeted moments.

Figure 6 shows two plots that illustrate the fit of our model with respect to disparities by income and discontinuities at rating thresholds. The plus signs in Panel (a) show the relationship between LSOA income and GP quality in the greater London subsample we use to estimate our model. This is analogous to Figure 1. We overlay this with red dots, which represent the same relationship as generated by our model estimates. The two are virtually identical, implying that our model effectively captures market shares across the income distribution in our data.

Figure 6
Model Fit: Matching Disparities and Regression Discontinuities



a. Quality Disparities by Income

b. Discontinuities

Notes: Panel (a) shows average GP quality by income computed from (i) the data and (ii) the baseline simulation from our model estimates. Panel (b) plots enrollment changes within rounded star rating bins both based on the data and implied by the baseline simulation from our model. All estimates are enrollment weighted.

Panel (b) of Figure 6 shows that our model is also able to effectively capture the *changes* in enrollment that we feature in our regression discontinuity. This shows that we obtain a good model fit beyond what would be mechanically captured by the large levels of inertia that we estimate. This chart repeats Figure 3, showing local linear smoothings within each rounded star rating for (i) observed enrollment changes in the greater London subsample (solid line) and (ii) enrollment changes simulated based on our parameter estimates (red dashed line). The two match closely. As in the data, our simulations show sharp responses at each star rating threshold and an upward slope across thresholds. Because we only explicitly target the *collapsed* version of these slopes and jumps, this panel shows the model's ability to match non-targeted moments.[31]

A natural concern is whether the GP fixed effects included in our model inadvertently capture the demand response at the rounding threshold, which is critical for identifying information precision in our model. In Appendix Figure D-1, we show the distribution of GP fixed effects is smooth across the rounding threshold indicating that the jump is primarily captured by the parameters governing information and preference for quality.

5.5 Counterfactual Simulations

We use our parameter estimates to construct counterfactual simulations and explore impacts of counterfactual policies that vary both information and access across the income distribution. Our main goal is to decompose the relative importance of information, access, and preferences in explaining the correlation between income and the quality of chosen GPs. While our regression discontinuity evidence and model results suggest that low income patients have relatively less precise information about quality, our counterfactual experiments shed light on whether information plays a quantitatively important role in these disparities.

Benchmark

As a benchmark, we construct a counterfactual simulation representing long-run choices for patients without access to information from star ratings. We focus on a benchmark without the star ratings to reflect the current status quo, since the star ratings were removed in early 2020. We construct long-run choices by setting $\varphi_{\ell t} = 1$, which reflects demand after choices have had time to adjust. In the absence of the star ratings, individuals have a prior based on the unconditional distribution of observed reviews in the data.

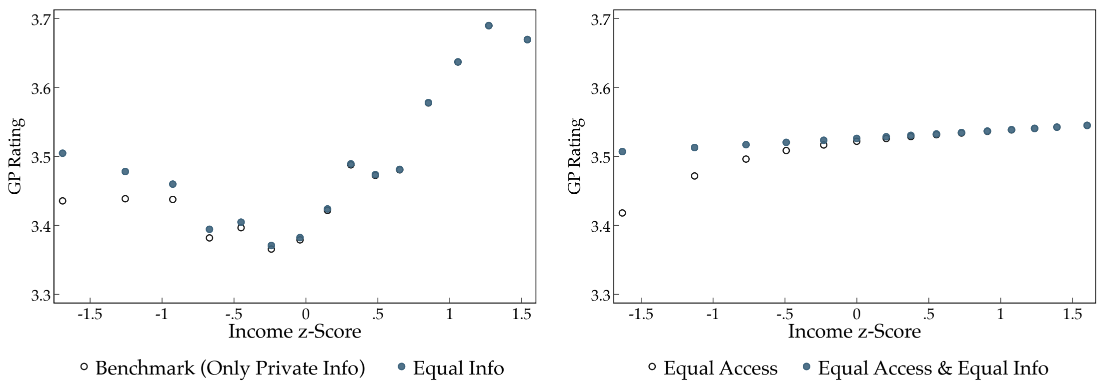
The hollow dots in Figure 7 panel (a) show the gradient between income and GP quality in this benchmark scenario. Without the website information, we still see a strong correlation between income and chosen quality, suggesting that higher income individuals continue to choose higher rated GPs.

One pattern to note is the presence of a relatively flat slope, or even slight negative slope, in bottom half of the distribution. This reflects the fact that, in the greater London sample,

[31]While the model fit is less good for GPs with ratings above 4.25, there is relatively little mass in this area.

many of the lowest income neighborhoods are located in relatively dense urban areas within relatively close proximity to high quality GPs.[32]

Figure 7
Relationship between Income and GP Rating
Under Counterfactual Information and Access to GPs



a. Equating Information

b. Equating Information and Access

Notes: Benchmark refers to no star rating system (only private information). Panel (a) shows the equal information counterfactual, i.e. giving all individuals the average private signal precision of the top 5% of the income distribution. Panel (b) shows counterfactuals with and without equal information when all individuals have the same choice set drawn randomly from the data. All counterfactual simulations examine the long-run (without inertia) averaged over the sample period.

The first row of Table 4 shows the correlation between income and quality under this benchmark, which is equal to 9.1%. We evaluate our other counterfactuals on the basis of the reduction in this correlation compared to the benchmark.

Equating Information

We now consider the impact of equating information across income groups (i.e., more equitable access to information). We ask what would happen if low income patients had the same precision of information as high income patients. We re-simulate our model setting the precision of private information for all individuals equal to the average precision in the top 5% of the income distribution. This allows us to determine the fraction of the observed income-quality gradient that can be explained by information disparities. As in the benchmark case, we consider the long-run ($\varphi_{\ell t} = 1$).

[32]This can be seen by noting that access for low income individuals in this sample is also high in Figure A-5.

Table 4
Summary of Counterfactual Information and Access

Counterfactual	Income-Quality Correlation	Percent Change Relative to Benchmark
Benchmark	0.091	
Equal Information	0.069	-24%
Equal Access	0.040	-55%
Equal Information + Equal Access	0.013	-86%
Stars	0.070	-22%
Stars + Equal Access	0.014	-85%

Notes: Benchmark refers to no star rating system and baseline choice sets for each individual. Equal information counterfactual gives all individuals the average private signal precision of the top 5% of the income distribution. Equal access gives all individuals the same choice set, which is drawn randomly from the data. Stars counterfactual allows individuals to incorporate the information from the star ratings. All counterfactual simulations examine the long-run ($\varphi_{\ell t} = 1$) averaged over the sample period. Percent change in correlation refers to the change in the unconditional correlation between LSOA income and r_{jt} between each counterfactual and the benchmark.

Equating information meaningfully reduces the income-quality gradient. The blue dots in panel (a) of Figure 7 show the gradient under the scenario. There is a notable jump in average quality chosen by lower income patients. As expected, the magnitude of this change declines across the income distribution. For the top half of the income distribution, there is no meaningful change in average quality of the chosen provider. The second row of Table 4 shows that equating information across income groups reduces the income quality correlation by 24% (from 9.1% to 6.9%). In sum, unequal information across income groups is an important driver of disparities in GP choice.

It is difficult to conceptualize a policy that would replicate this counterfactual. Real world informational treatments are unlikely to equate private signals across groups. As such, we next turn to examining the impact of an implementable policy that reduces the information gap between high and low income patients: the long run impact of the star rating system itself. We ask how large of a reduction in the income-quality gradient we would expect in the long run if the star ratings had continued to be displayed prominently on the website rather than being removed in 2020.

The information provided by star ratings reduces quality disparities nearly as much as providing equally precise private signals. This is seen in the 5th row of Table 4. Appendix Figure D-3 panel (a) shows that the star ratings lead to a large jump in average quality for low income patients, and small changes at higher levels of income. While the star ratings may not provide fine distinctions across practices, this suggests that the coarse information contained in star ratings helps equate choices.

One concern with these counterfactual simulations is that capacity constraints or wait times may limit the ability of patients to switch to higher quality providers. To account for this issue, we allow practitioners per patient to endogenously adjust in our counterfactual simulations.[33] The estimated coefficient on practitioners per patient is positive and significant, consistent with individuals having a distaste for wait times. However, allowing for endogenous patients per practitioner has a small effect on counterfactual estimates (see Appendix Table D-2).

Equating Access

How important is differential access to high quality GPs as a driver of disparities? We consider a counterfactual in which we provide all individuals with the same choice set. We implement this by randomly selecting 23 GP-LSOA-quarter observations in our data, where 23 is the median choices set size. We then created a simulated choice set based on the observable characteristics of each selected GP (e.g. quality or physician age) and the distance between the GP and LSOA. We assign this choice set to every LSOA, but hold fixed LSOA income. We then compute long-run market shares by assuming no inertia. We repeat this process 50 times and report the average across all repetitions.

Equating access reduces disparities to a greater degree than providing information alone, but substantial inequality remains. The hollow dots in panel (b) of Figure 7 show the gradient between income and quality after providing all patients the same choices. Given this uniform set of options, the relationship across the income distribution is relatively smooth in comparison to the patterns shown in panel (a). Overall, providing equal access reduces the correlation between income and quality by 55% relative to the benchmark (from 9.1% to 4.0%, as shown in Table 4).

Even with equal choice sets, there is a sizable gradient. While access appears to be a critical driver of differences in GP quality by income, roughly half of the gradient remains

[33] We allow patients to respond to changes in practitioners per patient where the number of patients is endogenously determined by predicted enrollment in the previous period under the counterfactual.

due to some combination of preferences and information.

Equating Access and Information

We now examine the combined effect of access and information to assess the fraction of the gradient that could be addressed through changes in the structure of the healthcare system and informational environment, and the fraction that results from patient preferences. First, we combine equal access with equal precision of consumers' private signals. Then, we consider combining equal access with the star ratings.

Equating both information and access substantially reduces disparities in GP quality. Table 4 shows that this combination of interventions reduces the income-quality gradient by nearly 90%, relative to the benchmark (from 9.1% to 1.3%). This also represents a reduction of over 65% compared to the scenario where we only equate access. The blue dots in panel (b) of Figure 7 show the gradient in this counterfactual, and a comparison with the hollow dots shows the incremental impact of information.[34]

The results in Table 4 show that the effect of equating both information and access is larger than the sum of its parts, i.e., equating information and access individually. The fact that the interaction is important reflects the idea that having access to a high quality GP is only valuable to individuals if they know about it.

Equating both information and access does not completely eliminate the relationship between income and GP quality. This is due to the remaining determinant of the relationship between income and GP quality, preferences. Indeed, we find a counterfactual correlation of 1.3% in this case. The estimates imply patients from high income LSOAs have (i) a slightly greater preference for quality and (ii) greater willingness (or ability) to travel to high-quality GPs. Our analysis is not targeted at understanding heterogeneity in preferences, and it may be that there is no role for policy to address these remaining differences. Still, the results imply that preferences play a relatively minor role after accounting for disparities in access and information.

---

[34] As in the previous counterfactual, access to the star rating system is able to provide much of the benefit of equating information precision despite the fact that the stars are a coarse signal (see Appendix Figure D-3 panel (b)).

6 Conclusion

The role of information as a driver of health care inequality has been understudied. We provide evidence that high income patients choose better GPs on average, in part due to the fact that these individuals have more precise private information. High income individuals may be better able to research GP quality or have greater access to recommendations from their social networks. In contrast, low income patients are significantly more responsive to online star rating thresholds, consistent with having less private information. These results are also in line with survey evidence that low income patients are more likely to use GP star ratings (Galizzi et al. 2012). Counterfactual simulations show that equating information leads to a significant reduction in the income-quality gradient. Access is also important, as is the interaction between access and information. This suggests that policies that combine information and access may be particularly cost effective.

An important caveat to these counterfactuals is the assumption of a fixed supply side. In our counterfactuals, we endogenize the number of practitioners per patient to account for crowding. However, we do not allow for endogenous entry, exit, or other long-run responses by GPs that might meaningfully alter choice sets. The consequences of such supply-side factors for disparities in healthcare, and for the income-quality gradient are ex-ante unclear, and a full model of the supply side is beyond the scope of this paper. However, our counterfactuals provide a benchmark from which to consider richer models that incorporate the incentives of GP practices.

Finally, there has been debate about allowing individuals the freedom to choose providers. In the UK, reforms have expanded choice for both GPs and hospitals.[35] While choice-based reforms can increase demand for high quality providers on average, such reforms may exacerbate inequality due to information gaps. Our findings highlight the importance of ensuring access to information when expanding choice.

[35] Choice of GPs was expanded in 2015, prior to our sample. In 2006, constraints on patients' choice of hospitals was relaxed. Gaynor et al. (2016) find that this raised hospital quality.

References

AHRQ, "National Healthcare Quality and Disparities Report," 2021.

Allende, Claudia, "Competition under social interactions and the design of education policies," *Working Paper*, 2022.

__ , Francisco Gallego, Christopher Neilson et al., "Approximating the equilibrium effects of informed school choice," Technical Report 2019.

Anderson, Michael and Jeremy Magruder, "Learning from the crowd: Regression discontinuity estimates of the effects of an online review database," *The Economic Journal*, 2012, *122* (563), 957–989.

Artmann, Elisabeth, Hessel Oosterbeek, and Bas van der Klaauw, "Do doctors improve the health care of their parents? Evidence from admission lotteries," *American Economic Journal: Applied Economics*, 2022, *14* (3), 164–84.

Balarajan, Yarlini, Selvaraj Selvaraj, and SV Subramanian, "Health care and equity in India," *The Lancet*, 2011, *377* (9764), 505–515.

Berry, Steven, James Levinsohn, and Ariel Pakes, "Differentiated products demand systems from a combination of micro and macro data: The new car market," *Journal of Political Economy*, 2004, *112* (1), 68–105.

Bettinger, Eric P, Bridget Terry Long, Philip Oreopoulos, and Lisa Sanbonmatsu, "The role of application assistance and information in college decisions: Results from the H&R Block FAFSA experiment," *The Quarterly Journal of Economics*, 2012, *127* (3), 1205–1242.

Bronnenberg, Bart J, Jean-Pierre Dubé, Matthew Gentzkow, and Jesse M Shapiro, "Do pharmacists buy Bayer? Informed shoppers and the brand premium," *The Quarterly Journal of Economics*, 2015, *130* (4), 1669–1726.

Brown, Zach Y, "An empirical model of price transparency and markups in health care," *Working Paper*, 2019.

__ and Jihye Jeon, "Endogenous information and simplifying insurance choice," *Working Paper*, 2022.

Calonico, Sebastian, Matias D Cattaneo, and Rocio Titiunik, "Robust data-driven inference in the regression-discontinuity design," *The Stata Journal*, 2014, *14* (4), 909–946.

__ , __ , Max H Farrell, and Rocio Titiunik, "Regression discontinuity designs using covariates," *Review of Economics and Statistics*, 2019, *101* (3), 442–451.

Cattaneo, Matias D, Michael Jansson, and Xinwei Ma, "Manipulation testing based on density discontinuity," *The Stata Journal*, 2018, *18* (1), 234–261.

__ , Nicolás Idrobo, and Rocío Titiunik, *A practical introduction to regression discontinuity designs: Foundations*, Cambridge University Press, 2019.

Chartock, Benjamin L, "Quality Disclosure, Demand, and Congestion: Evidence from Physician Ratings," 2021.

Chen, Yiqun, Petra Persson, and Maria Polyakova, "The roots of health inequality and the value of intra-family expertise," Technical Report, National Bureau of Economic Research 2019.

Chevalier, Judith A and Dina Mayzlin, "The effect of word of mouth on sales: Online book reviews," *Journal of Marketing Research*, 2006, *43* (3), 345–354.

Commission, Great Britain. Social Mobility, "State of the nation 2021: Social mobility and the pandemic," 2021.

Cookson, Richard, Carol Propper, Miqdad Asaria, and Rosalind Raine, "Socio-economic inequalities in health care in England," *Fiscal studies*, 2016, *37* (3-4), 371–403.

___ , Tim Doran, Miqdad Asaria, Indrani Gupta, and Fiorella Parra Mujica, "The inverse care law re-examined: a global perspective," *The Lancet*, 2021, *397* (10276), 828–838.

Cooper, Zack, Jr. Doyle Joseph J, John A Graves, and Jonathan Gruber, "Do Higher-Priced Hospitals Deliver Higher-Quality Care?," Working Paper 29809, National Bureau of Economic Research February 2022.

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

Dafny, Leemore and David Dranove, "Do report cards tell consumers anything they don't already know? The case of Medicare HMOs," *The RAND Journal of Economics*, 2008, *39* (3), 790–821.

Darden, Michael and Ian M McCarthy, "The star treatment estimating the impact of star ratings on Medicare Advantage enrollments," *Journal of Human Resources*, 2015, *50* (4), 980–1008.

Devaux, Marion, "Income-related inequalities and inequities in health care services utilisation in 18 selected OECD countries," *The European Journal of Health Economics*, 2015, *16* (1), 21–33.

Doyle, Joseph, John Graves, and Jonathan Gruber, "Evaluating measures of hospital quality: Evidence from ambulance referral patterns," *Review of Economics and Statistics*, 2019, *101* (5), 841–852.

Dranove, David, Daniel Kessler, Mark McClellan, and Mark Satterthwaite, "Is More Information Better? The Effects of Report Cards on Health Care Providers," *Journal of Political Economy*, June 2003, *111* (3), 555–588.

Fu, Chao and Jesse Gregory, "Estimation of an equilibrium model with externalities: Post-disaster neighborhood rebuilding," *Econometrica*, 2019, *87* (2), 387–421.

Galizzi, Matteo Maria, Marisa Miraldo, Charitini Stavropoulou, Mihir Desai, Wikum Jayatunga, Mitesh Joshi, and Sunny Parikh, "Who is more likely to use doctor-rating websites, and why? A cross-sectional study in London," *BMJ open*, 2012, *2* (6), e001493.

Gaynor, Martin, Carol Propper, and Stephan Seiler, "Free to choose? Reform, choice, and consideration sets in the English National Health Service," *American Economic Review*, 2016, *106* (11), 3521–57.

Gourieroux, C., A. Monfort, and E. Renault, "Indirect Inference," *Journal of Applied Econometrics*, 1993, *8*, S85–S118.

Greaves, Felix, Utz J Pape, Dominic King, Ara Darzi, Azeem Majeed, Robert M Wachter, and Christopher Millett, "Associations between internet-based patient ratings and conventional surveys of patient experience in the English NHS: an observational study," *BMJ Quality & Safety*, 2012, *21* (7), 600–605.

___ , Utz J. Pape, Dominic King, Ara Darzi, Azeem Majeed, Robert M. Wachter, and Christopher Millett, "Associations Between Web-Based Patient Ratings and Objective Measures of Hospital Quality," *Archives of Internal Medicine*, 03 2012, *172* (5), 435–436.

Grennan, Matthew and Robert J Town, "Regulating innovation with uncertain quality: information, risk, and access in medical devices," *American Economic Review*, 2020, *110* (1), 120–161.

Gwatkin, Davidson R, Abbas Bhuiya, and Cesar G Victora, "Making health systems more equitable," *The Lancet*, 2004, *364* (9441), 1273–1280.

Handel, Benjamin R, Jonathan T Kolstad, Thomas Minten, and Johannes Spinnewijn, "The social determinants of choice quality: evidence from health insurance in the Netherlands," Technical Report, National Bureau of Economic Research 2020.

Hart, Julian Tudor, "The inverse care law," *The Lancet*, 1971, *297* (7696), 405–412.

Hastings, Justine, Christopher A Neilson, and Seth D Zimmerman, "The effects of earnings disclosure on college enrollment decisions," Technical Report, National Bureau of Economic Research 2015.

Hastings, Justine S and Jeffrey M Weinstein, "Information, school choice, and academic achievement: Evidence from two experiments," *The Quarterly Journal of Economics*, 2008, *123* (4), 1373–1414.

Hull, Peter, "Estimating hospital quality with quasi-experimental data," *Available at SSRN 3118358*, 2018.

Imbens, Guido and Karthik Kalyanaraman, "Optimal Bandwidth Choice for the Regression Discontinuity Estimator," *The Review of Economic Studies*, 2012, *79* (3), 933–959.

Jin, Ginger Zhe and Alan T Sorensen, "Information and consumer choice: the value of publicized health plan ratings," *Journal of Health Economics*, 2006, *25* (2), 248–275.

Kapor, Adam J, Christopher A Neilson, and Seth D Zimmerman, "Heterogeneous beliefs and school choice mechanisms," *American Economic Review*, 2020, *110* (5), 1274–1315.

Kolstad, Jonathan T, "Information and quality when motivation is intrinsic: Evidence from surgeon report cards," *American Economic Review*, 2013, *103* (7), 2875–2910.

_ **and Michael E Chernew**, "Quality and consumer decision making in the market for health insurance and health care services," *Medical Care Research and Review*, 2009, *66* (1), 28S–52S.

Kowalski, Radoslaw, "Patients' written reviews as a resource for public healthcare management in England," *Procedia Computer Science*, 2017, *113*, 545–550.

Kummer, Michael E, Ulrich Laitenberger, Cyrus E Rich, Danny R Hughes, and Turgay Ayer, "Healthy reviews! Online physician ratings reduce healthcare interruptions," Technical Report, ZEW Discussion Papers 2021.

Lewis, Gregory and Georgios Zervas, "The welfare impact of consumer reviews: A case study of the hotel industry," Technical Report, Working Paper 2016.

Luca, Michael, "Reviews, reputation, and revenue: The case of Yelp. com," 2016. Harvard Business School Working Paper.

_ **and Georgios Zervas**, "Fake it till you make it: Reputation, competition, and Yelp review fraud," *Management Science*, 2016, *62* (12), 3412–3427.

Marmot, Michael, Commission on Social Determinants of Health et al., "Achieving health equity: from root causes to fair outcomes," *The Lancet*, 2007, *370* (9593), 1153–1163.

Mayzlin, Dina, Yaniv Dover, and Judith Chevalier, "Promotional reviews: An empirical investigation of online review manipulation," *American Economic Review*, 2014, *104* (8), 2421–55.

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

Newberry, Peter and Xiaolu Zhou, "Heterogeneous effects of online reputation for local and national retailers," *International Economic Review*, 2019, *60* (4), 1565–1587.

NHS, "NHS England response to the specific duties of the Equality Act Equality information relating to public facing functions," 2016.

O'Dowd, Adrian, "Poverty status is linked to worse quality of care," *British Medical Journal*, 2020.

OECD, "Health at a Glance: Europe 2014," 2014.

Peters, David H, Anu Garg, Gerry Bloom, Damian G Walker, William R Brieger, and M Hafizur Rahman, "Poverty and access to health care in developing countries," *Annals of the new York Academy of Sciences*, 2008, *1136* (1), 161–171.

Petrin, Amil, "Quantifying the benefits of new products: The case of the minivan," *Journal of Political Economy*, 2002, *110* (4), 705–729.

Pope, Devin G, "Reacting to rankings: evidence from "America's Best Hospitals"," *Journal of Health Economics*, 2009, *28* (6), 1154–1165.

Raval, Devesh and Ted Rosenbaum, "Why do previous choices matter for hospital demand? decomposing switching costs from unobserved preferences," *Review of Economics and Statistics*, 2018, *100* (5), 906–915.

Reimers, Imke and Joel Waldfogel, "Digitization and pre-purchase information: the causal and welfare impacts of reviews and crowd ratings," *American Economic Review*, 2021, *111* (6), 1944–71.

Santos, Rita, Hugh Gravelle, and Carol Propper, "Does quality affect patients' choice of doctor? Evidence from England," *The Economic Journal*, 2017, *127* (600), 445–494.

Scobie, Sarah and Jessica Morris, "Quality and inequality: digging deeper," Technical Report, Nuffield Trust January 2020.

Shepard, Mark, "Hospital network competition and adverse selection: evidence from the Massachusetts health insurance exchange," *American Economic Review*, 2022, *112* (2), 578–615.

van Doorslaer, Eddy, Cristina Masseria, Xander Koolman et al., "Inequalities in access to medical care by income in developed countries," *Cmaj*, 2006, *174* (2), 177–183.

_, Xander Koolman, and Andrew M Jones, "Explaining income-related inequalities in doctor utilisation in Europe," *Health economics*, 2004, *13* (7), 629–647.

Werner, Rachel M, Edward C Norton, R Tamara Konetzka, and Daniel Polsky, "Do consumers respond to publicly reported quality information? Evidence from nursing homes," *Journal of Health Economics*, 2012, *31* (1), 50–61.

Xiao, Mo, "Is quality accreditation effective? Evidence from the childcare market," *International Journal of Industrial Organization*, 2010, *28* (6), 708–721.

Online Appendix

A Additional Summary Statistics and Background

A.1 Supplemental Data Appendix

We supplement the primary data on GP practice reviews and enrollment with the following additional data.

GP Patient Survey The GP Patient Survey is an independently-run representative annual survey of over 1 million individuals that is run on behalf of NHS England.[36] The survey was conducted twice a year from July 2011 to March 2016, and after that point was conducted annually. We match this to quarterly data using the closest available survey date.

Quality and Outcomes Framework Score The Quality and Outcomes Framework (QOF) is a system used for performance pay of GPs. We focus on the two overall scores. The clinical score aggregates a number of clinical indicators, such as whether a GP provided proper vaccinations and performed necessary tests for patients with specific diagnoses. We also examine the overall score, which includes indicators such as whether proper training was provided to GP staff. These scores are available online but it is relatively difficult for patients to compare scores across GP practices.

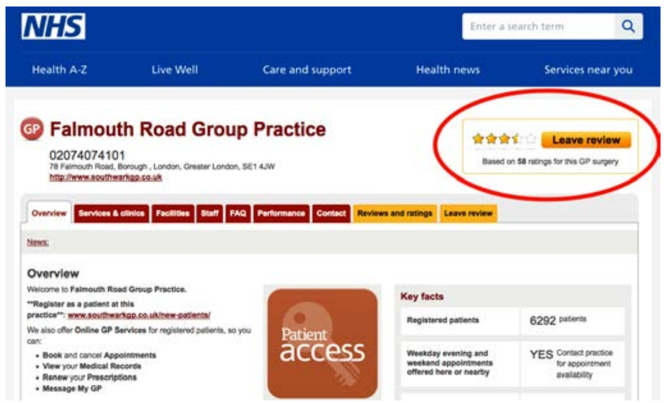
GP Characteristics GP practice characteristics were obtained from NHS Digital Organisation Data Service. The location of each GP was obtained by geocoding the address of each GP in the "GP Practices" file. The opening date was also obtained from this file. The number of practitioners in a GP practice and each practitioner's experience was obtained from the "GP Practitioners" file.

[36]The survey data is available at `https://www.gp-patient.co.uk/`.

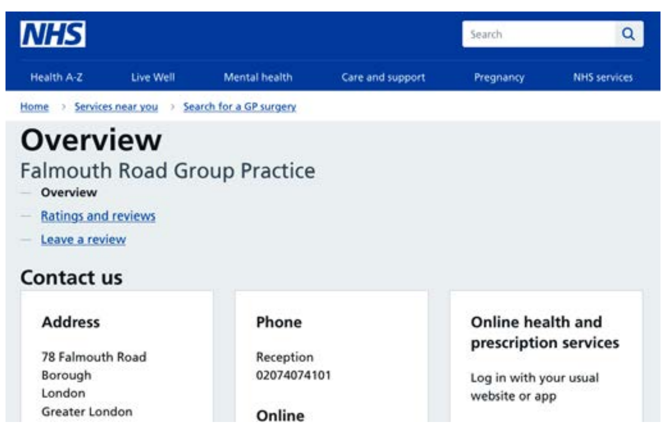
A.2 Descriptive Figures and Tables

Figure A-1
Examples of the NHS Website

Panel A: With Visible Star Rating (Prior to 2020)

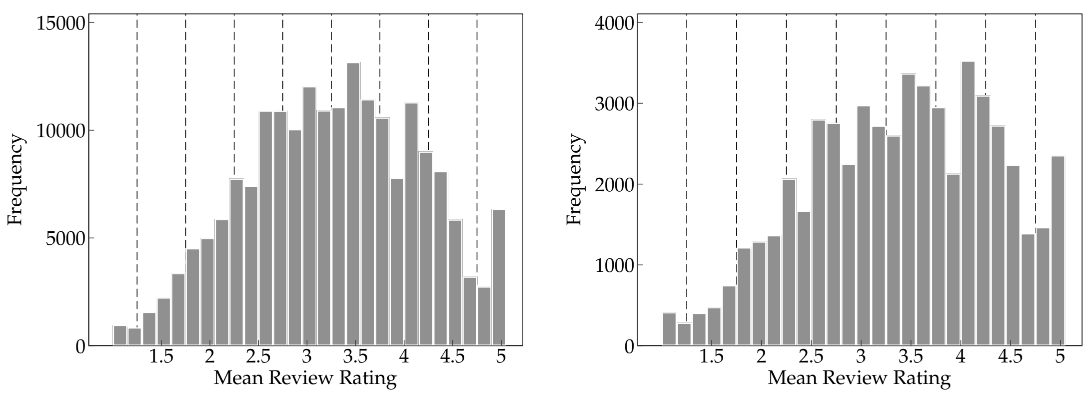


Panel B: Without Visible Star Rating (2020 and After)



Notes: An example of the NHS website for a single GP practice prior to January 2020 (with visible star ratings) and after January 2020 (when the star ratings were removed).

Figure A-2
Histogram of Average GP Reviews



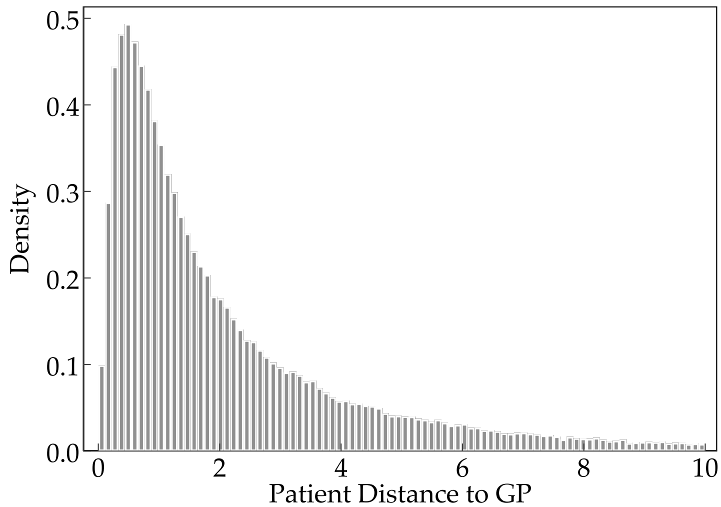
a. Visible star ratings

b. No star ratings

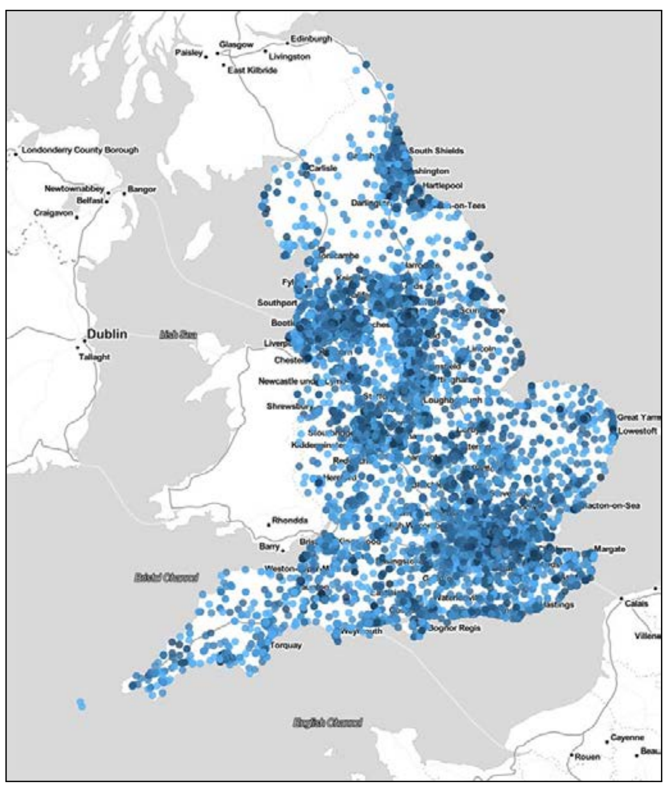
Notes: Chart shows histogram of average reviews r_{jt} for the period when the NHS Choices website displayed star ratings (“visible star ratings”) and the period when the website did not display star ratings (“no star ratings”). The NHS calculated average reviews using the running average of individual reviews over the previous two years. Vertical lines show thresholds for rounding star ratings.

Figure A-3
Travel Distance to GPs

Panel A: Histogram of Distance to Chosen GP

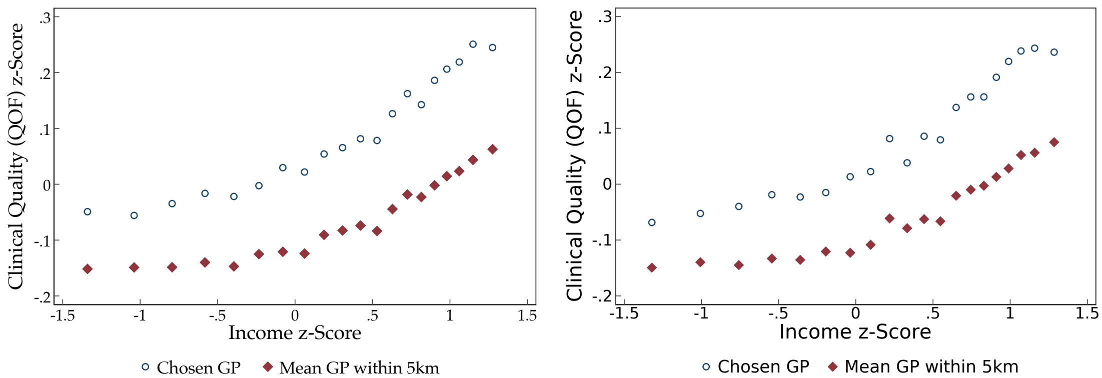


Panel B: GP Location and Enrollment



Notes: Panel A shows a histogram of distance between each individual's LSOA centroid and their chosen GP. Panel B shows the location of all GPs in England. Darker colors correspond to higher enrollment.

Figure A-4
Relationship between Income and Objective Measure of GP Quality

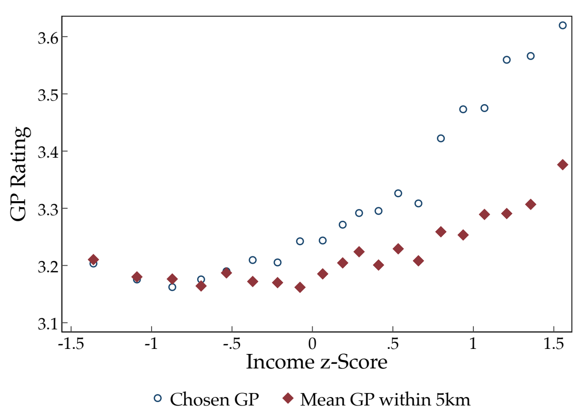


a. All Patients

b. Movers Only

Notes: Binscatter plots show relationship between LSOA income and the quality of patients' chosen GP (in blue) as well as the mean quality of GPs within 5km (in red). Quality is measured using the Quality Outcome Framework (QOF) clinical quality score, an objective measure of GP quality used for performance pay. Panel (a) includes all patients and Panel (b) includes the sample of patients who moved residences and chose a new GP. The unit of observation is a LSOA-GP-quarter and results are weighted by the quarterly change in enrollment in Panel (a) and enrollment in Panel (b).

Figure A-5
Relationship between Income and GP Rating
Greater London



Notes: Binscatter plots show relationship between LSOA income and the quality of patients' chosen GP (in blue) as well as the mean quality of GPs within 5km (in red). Quality is measured as the average patient rating on the NHS website. The unit of observation is an LSOA-GP-quarter and results are weighted by the quarterly change in enrollment.

Table A-1

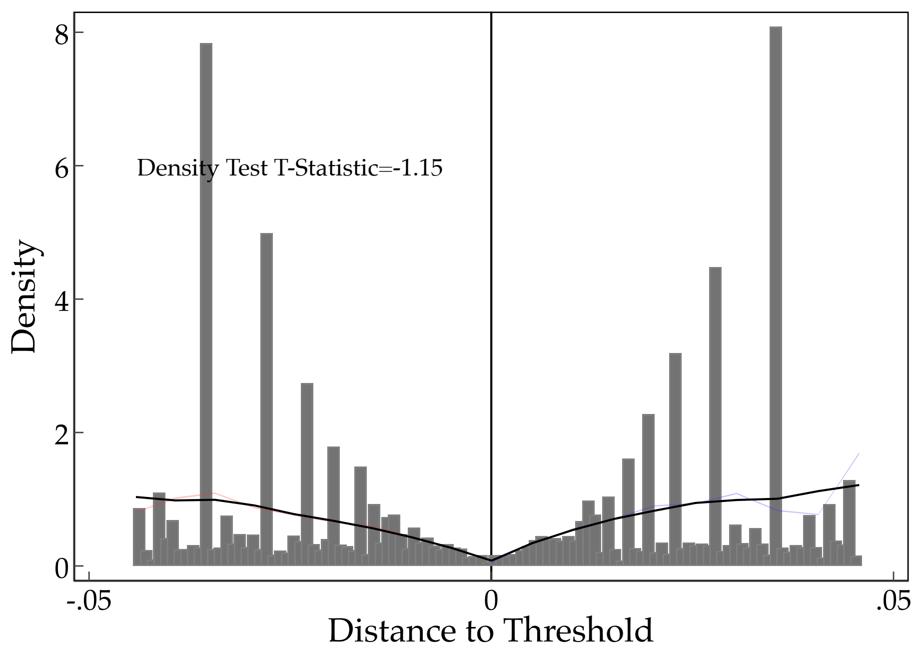
Correlation of Subjective Reviews with Other Quality Measures

	All		< 5 Reviews		≥ 5 Reviews	
	Corr	p-value	Corr	p-value	Corr	p-value
Patient Surveys:						
Easy getting through to GP	0.45	0.000	0.31	0.000	0.48	0.000
Receptionist was helpful	0.44	0.000	0.32	0.000	0.46	0.000
Able to get appointment	0.45	0.000	0.35	0.000	0.47	0.000
GP gave enough time	0.43	0.000	0.35	0.000	0.43	0.000
GP explained well	0.39	0.000	0.31	0.000	0.40	0.000
GP involved you	0.41	0.000	0.33	0.000	0.41	0.000
GP treated you with care and concern	0.42	0.000	0.34	0.000	0.43	0.000
Confidence and trust in GP	0.37	0.000	0.30	0.000	0.38	0.000
Overall experience good	0.52	0.000	0.41	0.000	0.55	0.000
Quality and Outcomes Framework:						
Clinical (z-score)	0.17	0.000	0.13	0.000	0.20	0.000
Overall (z-score)	0.16	0.000	0.13	0.000	0.19	0.000
Prescription Drugs:						
Prescriptions per Patient	-0.00	0.938	-0.00	0.493	-0.04	0.000
Addictive Prescriptions per Patient	0.02	0.000	0.02	0.010	0.01	0.035

Notes: Correlation coefficients between quality measures and variables and mean patient review r_{jt} along with the relevant p-value. < 5 and ≥ 5 refer to subsamples based on the quantity of reviews left for each GP.

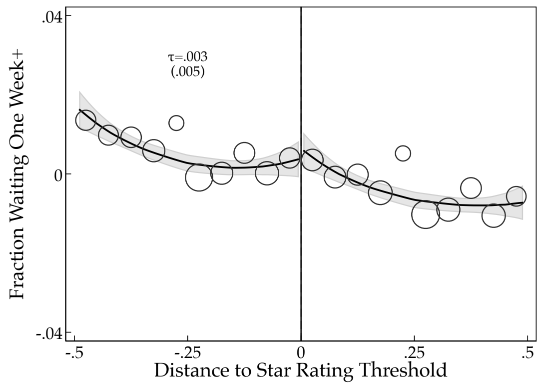
B Supplementary Analysis for Regression Discontinuity Approach

Figure B-1
Density Tests of GP Average Reviews Around Star Rating
Thresholds

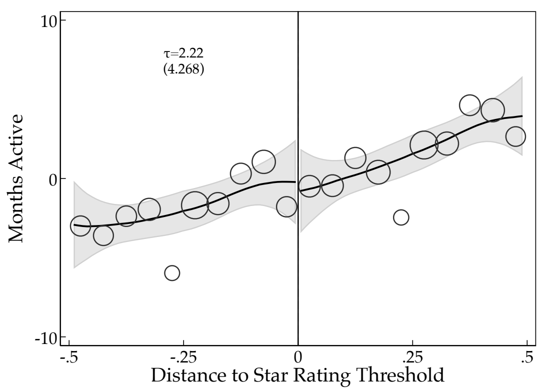


Notes: The plot shows a histogram and polynomial density estimate for practices above and below star rounding thresholds using our RD analysis sample. The unrestricted robust bias-corrected t-statistic following Cattaneo et al. (2018) is shown in the upper left hand corner.

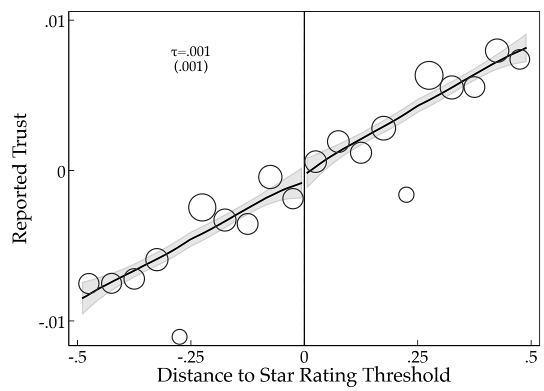
Figure B-2
Smoothness of GP-level Covariates Around Threshold



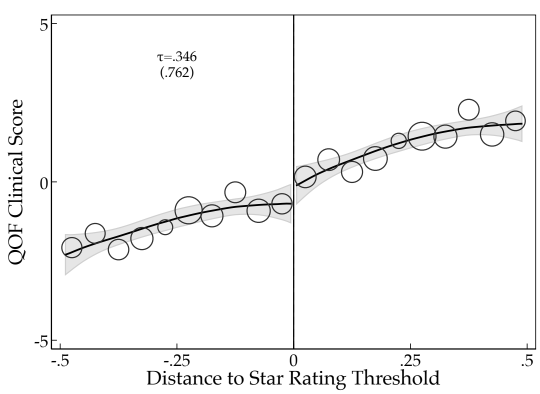
a. Waiting Times



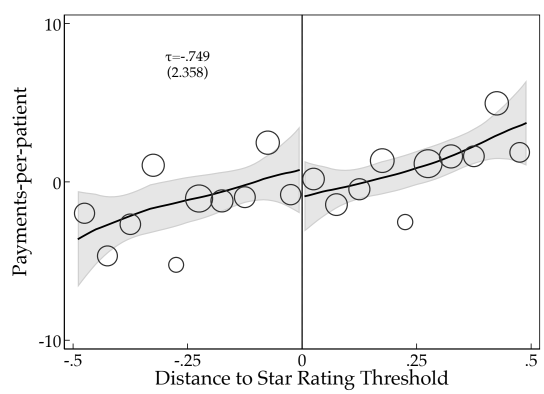
b. Months Active



c. Confidence and Trust (Survey)



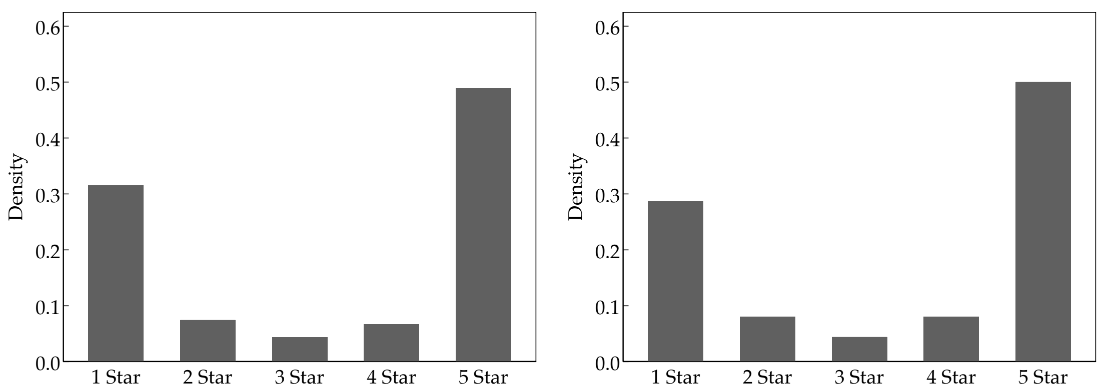
d. QOF Clinical Score



e. Payments Per-Patient

Notes: Changes at the rounding threshold for GP-level observables. Circle size corresponds to the number of observations in each bin. Fitted lines are from a local linear regressions with a triangular kernel. Shaded area shows 95% confidence intervals. Here τ refers to the RD point estimate and standard errors are shown in parenthesis.

Figure B-3
Distribution of Individual Reviews Around Website Change

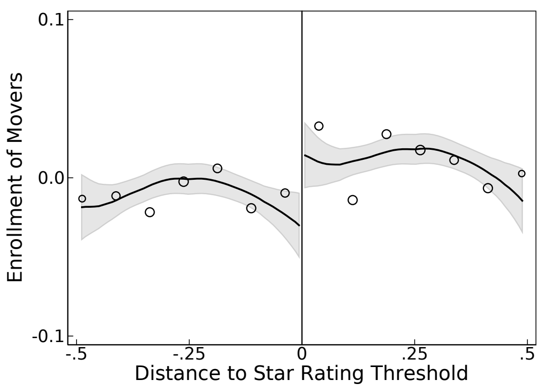


a. Visible Star Ratings

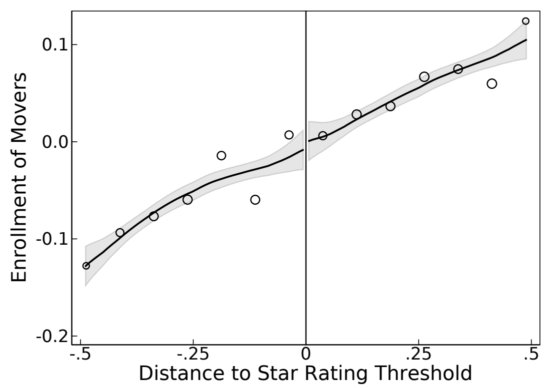
b. No Star Ratings

Notes: Distribution of individual reviews a month before and a month after the website change in January 2020. We do not find a statistically significant difference in the distributions. The Chi-squared test statistic is 5.6556 (p-value 0.226).

Figure B-4
Effect of Star Rating Threshold on GP Enrollment by Income
Sample of Movers



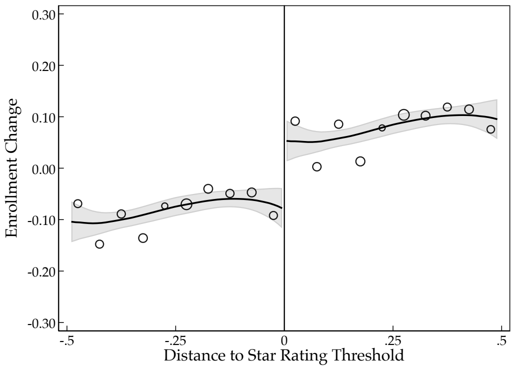
a. Low Income



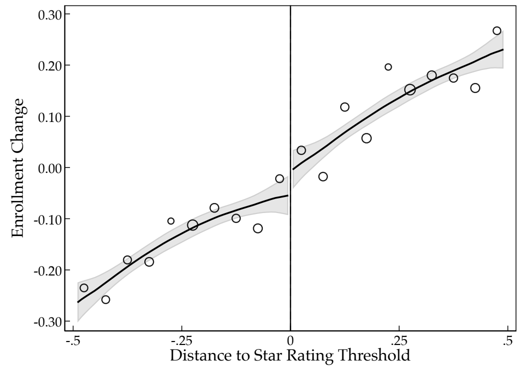
b. High Income

Notes: Chart shows mean enrollment change around the threshold for star ratings for movers, splitting at the median of LSOA income, for the period in which star ratings were visible. Movers are defined as individuals who relocated to a new LSOA and registered with a new GP practice in the same month. The size of the circles correspond to the number of observations in each bin. The fitted line is from a local linear regression using a triangular kernel. Shaded area shows 95% confidence interval.

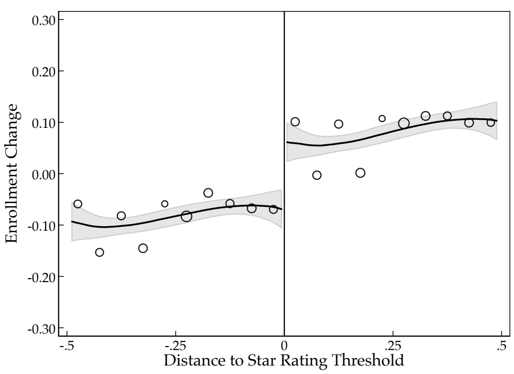
Figure B-5
Effect of Star Rating Threshold on Enrollment Change
Additional Heterogeneity Analysis



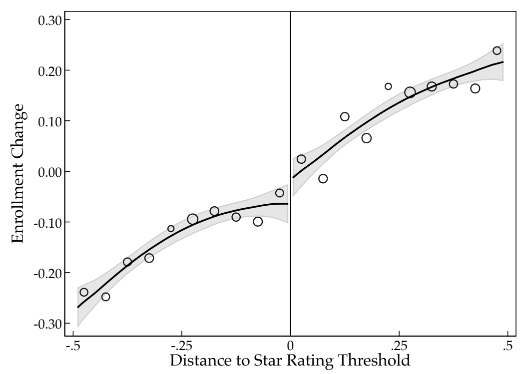
a. Low Education



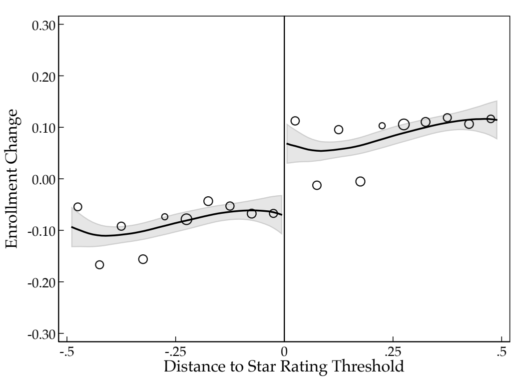
b. High Education



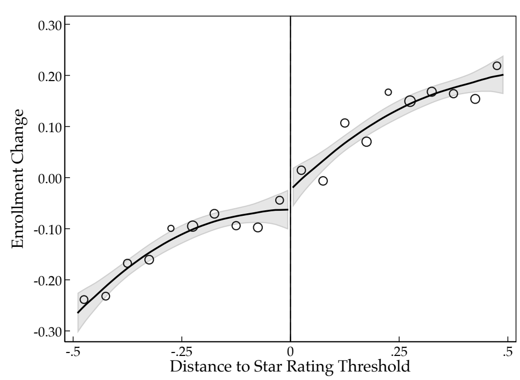
c. Low Employment



d. High Employment



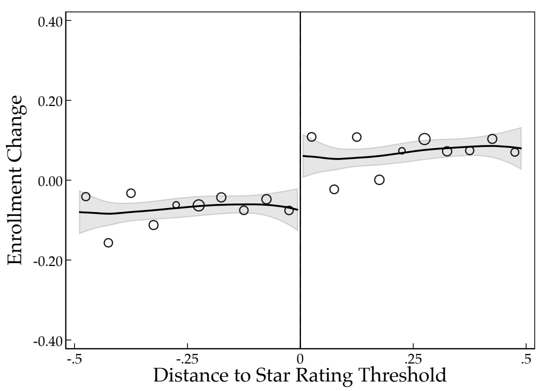
e. Low Health



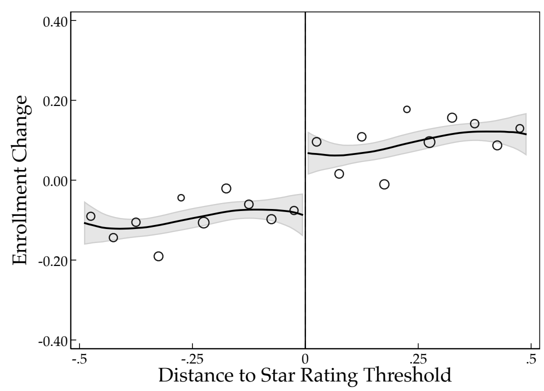
f. High Health

Notes: Mean enrollment change around threshold for star ratings, splitting at the median of LSOA level education, employment, and health status. The size of the circles correspond to the number of observations in each bin. Fitted lines are from local linear regressions using a triangular kernel. Shaded area shows 95% confidence interval.

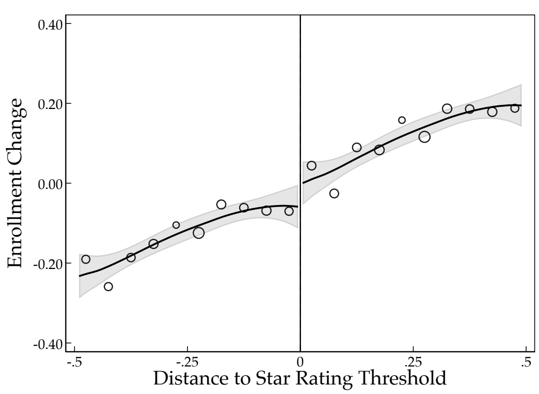
Figure B-6
Effect of Star Rating Threshold on GP Enrollment
by Income Quartile



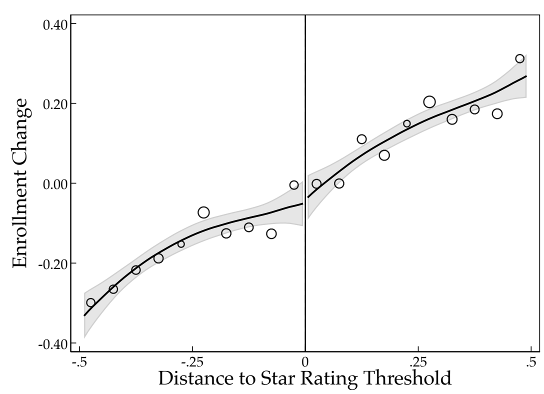
a. First Quartile
(Lowest Income)



b. Second Quartile



c. Third Quartile



d. Fourth Quartile
(Highest Income)

Notes: Mean enrollment change around threshold for star ratings by quartile of LSOA income in period in which star ratings were visible. The size of the circles correspond to the number of observations in each bin. Fitted lines are from local linear regressions using a triangular kernel. Shaded area shows 95% confidence interval.

Table B-1
RD Estimates for Full Sample
Robustness to Alternative Specifications and Bandwidths

	Alternative Specifications				
	Epanechnikov Kernel	No Min. # Reviews	Include $r_{jt} = c_s$	No Covariates	% Change Enroll.
Estimate	0.135**	0.112**	0.097**	0.082	0.276***
	(0.060)	(0.052)	(0.049)	(0.051)	(0.100)
Bandwidth	0.12	0.15	0.15	0.17	0.12
N	824,685	1,214,654	1,077,599	1,194,966	846,362
	Alternative Bandwidths				
	Bandwidth=0.1	Bandwidth=0.2	Bandwidth=0.3	Bandwidth=0.4	Bandwidth=0.5
Estimate	0.157**	0.110**	0.086**	0.072**	0.067**
	(0.067)	(0.045)	(0.037)	(0.034)	(0.031)
Bandwidth	0.10	0.20	0.30	0.40	0.50
N	698,624	1,431,288	2,168,005	2,877,100	3,517,643

Notes: Panel (a) shows robustness for RD specifications. The first column employs the epanechnikov kernel, the second includes GP practices with fewer than 5 reviews, the third includes observations with the index r_{jt} exactly equal to a rounding threshold, the fourth excludes all covariates, and the fifth uses the percent change in enrollment at the LSOA-GP level as the dependent variable. Panel (b) shows robustness to bandwidth choice. Sample period is when stars were visible. Except where otherwise noted, the dependent variable is quarterly enrollment change for an LSOA-GP, and controls for GP age, age squared, and number of practitioners in the GP practice, as well as threshold fixed effects are included. Unless otherwise noted, we follow Calonico et al. (2019) and Calonico et al. (2014) to select bandwidths, include covariates and, calculate standard errors clustered at the GP level (in parentheses). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B-2
RD Estimates by Income
Robustness to Alternative Specifications and Bandwidths

	Panel a: Alternative Specifications									
	Epanechnikov Kernel		No Min. # Reviews		Include $r_{jt} = c_s$		No Covariates		% Change Enroll.	
	Low Inc.	High Inc.	Low Inc.	High Inc.	Low Inc.	High Inc.	Low Inc.	High Inc.	Low Inc.	High Inc.
Estimate	0.213***	0.059	0.185***	0.057	0.166***	0.035	0.161**	0.005	0.463***	0.063
	(0.073)	(0.074)	(0.068)	(0.071)	(0.058)	(0.062)	(0.067)	(0.060)	(0.158)	(0.065)
Bandwidth	0.12	0.11	0.15	0.13	0.17	0.13	0.15	0.18	0.12	0.16
N	420,139	397,821	547,867	498,345	634,003	499,627	560,213	636,077	419,931	591,464
	Panel b: Movers and Alternative Bandwidths									
	Movers Only		Bandwidth=0.1		Bandwidth=0.3		Bandwidth=0.4		Bandwidth=0.5	
	Low Inc.	High Inc.	Low Inc.	High Inc.	Low Inc.	High Inc.	Low Inc.	High Inc.	Low Inc.	High Inc.
Estimate	0.118**	0.000	0.223***	0.095	0.145***	0.027	0.119***	0.027	0.111***	0.024
	(0.059)	(0.059)	(0.084)	(0.082)	(0.047)	(0.045)	(0.042)	(0.040)	(0.038)	(0.037)
Bandwidth	0.10	0.11	0.10	0.10	0.30	0.30	0.40	0.40	0.50	0.50
N	336,678	391,313	347,252	351,372	1,076,997	1,091,008	1,424,869	1,452,231	1,742,940	1,774,703

Notes: Panel (a) shows robustness for RD specifications. The first set of columns employs the epanechnikov kernel, the second includes GP practices with fewer than 5 reviews, the third includes observations with the index r_{jt} exactly equal to a rounding threshold, the fourth excludes all covariates, and the fifth uses the percent change in enrollment at the LSOA-GP level as the dependent variable. Panel (b) shows robustness to limiting to movers and the bandwidth choice. Movers are defined as individuals who relocated to a new LSOA and registered with a new GP practice in the same month. Sample period is when stars were visible, and standard errors clustered at the GP level are included in parenthesis. Except where otherwise noted, the dependent variable is quarterly enrollment change for an LSOA-GP, and controls for GP age, age squared, and number of practitioners in the GP practice, as well as threshold fixed effects are included. Unless otherwise noted, we follow Calonico et al. (2019) and Calonico et al. (2014) to select bandwidths, include covariates, and calculate standard errors clustered at the GP level (in parentheses). High and low income refer to above versus below median income deprivation at the LSOA level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B-3
RD Estimates by Income and Age
Robustness Addressing Heterogenous Website Usage by Age

	Young LSOAs		Old LSOAs	
	Low Income	High Income	Low Income	High Income
Estimate	0.210**	0.128	0.141	0.019
	(0.082)	(0.126)	(0.087)	(0.066)
Bandwidth	0.15	0.11	0.14	0.17
N	1,137,669	674,103	605,271	1,100,600

Notes: Dependent variable is change in GP enrollment. The first two columns limit the sample to LSOAs where the fraction of individuals ages 20-44 is above the median. The last two columns limit the sample to LSOAs where the fraction of individuals ages 20-44 is below the median. The dependent variable is quarterly enrollment change for an LSOA-GP. Controls for GP age, age squared, and number of practitioners in the GP practice, as well as threshold fixed effects are included. We follow Calonico et al. (2019) and Calonico et al. (2014) to select bandwidths, include covariates, calculate standard errors clustered at the GP level (in parentheses). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B-4
RD Estimates by Income
Robustness Addressing Potential Capacity Constraints

	Similar Choice Set		No GPs with Static Enrollment	
	Low Inc.	High Inc.	Low Inc.	High Inc.
Estimate	0.159**	0.100	0.191***	0.077
	(0.075)	(0.083)	(0.072)	(0.078)
Bandwidth	0.14	0.12	0.15	0.13
N	363,094	285,300	492,612	403,975

Notes: The first two columns limit the sample to high (low) income LSOAs that are within 1km of a low (high) income LSOA, implying that low and high income individuals face a similar choice set. Columns 3 and 4 remove GPs with enrollment that does not change between at least two consecutive months within a year. The dependent variable is quarterly enrollment change for an LSOA-GP. Controls for GP age, age squared, and number of practitioners in the GP practice, as well as threshold fixed effects are included. We follow Calonico et al. (2019) and Calonico et al. (2014) to select bandwidths, include covariates, calculate standard errors clustered at the GP level (in parentheses). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B-5
Effect of Star Ratings on Enrollment Change
Simple Parametric Regression Discontinuity Estimates

	Full Sample		Structural Sample	
	Low Income	High Income	Low Income	High Income
Estimate	0.110***	0.027	0.150**	0.040
	(0.034)	(0.033)	(0.067)	(0.062)
Distance from threshold	0.115	0.500***	0.110	0.590***
	(0.073)	(0.070)	(0.143)	(0.135)
N	1,698,686	1,722,858	481,330	513,175

Notes: Results from Parametric linear RD specification with a single slope coefficient above and below each threshold. The dependent variable is the quarterly change in GP-LSOA enrollment. "Full Sample" includes all observations meeting the criteria for the regressions shown in Table 2 for the period with visible star ratings, but using a bandwidth of 0.5. "Structural Sample" includes observations that meet the same criteria as the Full Sample, but limits to Greater London and LSOA-GP pairs that meet the choice set criteria described in Section 5.2. Each regression controls cutoff-specific fixed effects for each rounding threshold. Includes star ratings 1.5 to 4.5. Standard errors are in parentheses and clustered by GP. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

C Panel Fixed Effects Regressions

In addition to using a regression discontinuity approach, we examine whether enrollment responds to *within*-GP variation in rounded star ratings. Because individual reviews are added and subtracted each quarter, a GP practice's star rating may differ over time even if underlying quality remains unchanged. As a result, we are able to estimate two-way fixed effects specifications that account for all time-invariant practice level factors. Specifically, we estimate:

$$y_{j\ell t} = \gamma s_{jt} + X'_{jt}\beta + \alpha_j + \delta_t + \epsilon_{j\ell t} \tag{C-1}$$

where s_{jt} is the rounded star rating for GP j in quarter t, α_j are GP fixed effects, and δ_t are quarter-year fixed effects.[37] We also control for time-varying characteristics of the GP, X_{jt}, including age of the GP practice, age squared, and the number of practitioners in the GP practice. The main coefficient of interest is γ, the impact of an increase in rounded star ratings. As in our baseline RD specification, the outcome of interest is the quarterly change in enrollment at the GP-LSOA level and standard errors are clustered at the GP practice level.

The results, presented in Appendix Table C-1, are consistent with our RD strategy. In the full sample, star ratings have a positive and significant effect on quarterly change in enrollment. In Column 2 of Appendix Table C-1 we interact s_{jt} with an indicator equal to one if the LSOA has below median income. In line with our earlier results, we find that low income individuals respond more to the star ratings. In Columns 3 and 4, we estimate a non-parametric version of the regression, allowing the coefficients to vary for each value of the star rating and star ratings interacted with income. The results indicate that demand is monotonically increasing in star ratings, and confirm that low income areas are more responsive.

These findings provide additional evidence that star ratings effect demand, particularly for low income individuals. This is largely consistent with the RD estimates, although point estimates tend to be somewhat smaller in magnitude using the two-way fixed effect ap-

[37] For ease of interpretation, we include $2 \times s_{jt}$ in practice. The coefficient magnitudes can therefore be viewed as the impact of a one step increase in the ratings.

proach. Interpreting these estimates as the causal effect of the rounded star ratings requires a slightly stronger set of assumptions—namely that the fixed effects capture any underlying practice-quality that might drive patient demand (absent ratings), and hence that changes in star ratings are not correlated with any other changes in patients information about practice quality.

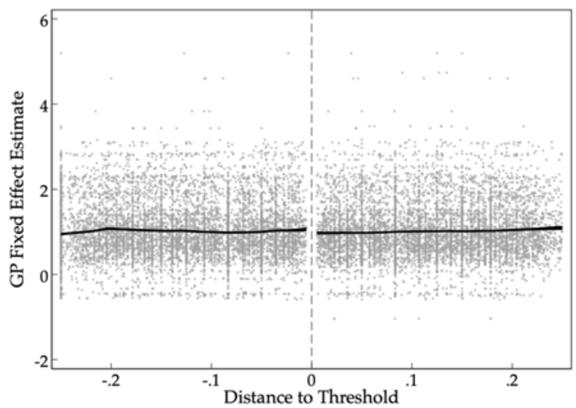
Table C-1
Effect of Star Ratings on Enrollment Change
Panel Regression Estimates

	(1)	(2)	(3)	(4)
Stars × 2	0.029∗∗∗	0.025∗∗∗		
	(0.001)	(0.001)		
(Stars×2) × 1(Low Income)		0.008∗∗∗		
		(0.001)		
1(Stars=1.5)			−0.034∗	−0.103∗∗∗
			(0.017)	(0.018)
1(Stars=2)			0.016	−0.054∗∗
			(0.016)	(0.017)
1(Stars=2.5)			0.045∗∗	−0.017
			(0.017)	(0.017)
1(Stars=3)			0.060∗∗∗	0.014
			(0.017)	(0.018)
1(Stars=3.5)			0.095∗∗∗	0.063∗∗∗
			(0.017)	(0.017)
1(Stars=4)			0.125∗∗∗	0.101∗∗∗
			(0.017)	(0.017)
1(Stars=4.5)			0.160∗∗∗	0.145∗∗∗
			(0.018)	(0.018)
1(Stars=5)			0.185∗∗∗	0.166∗∗∗
			(0.019)	(0.020)
1(Stars=1.5)× 1(Low Income)				0.113∗∗∗
				(0.012)
1(Stars=2)× 1(Low Income)				0.115∗∗∗
				(0.007)
1(Stars=2.5)× 1(Low Income)				0.100∗∗∗
				(0.006)
1(Stars=3)× 1(Low Income)				0.073∗∗∗
				(0.008)
1(Stars=3.5)× 1(Low Income)				0.045∗∗∗
				(0.005)
1(Stars=4)× 1(Low Income)				0.029∗∗∗
				(0.006)
1(Stars=4.5)× 1(Low Income)				0.006
				(0.008)
1(Stars=5)× 1(Low Income)				0.015
				(0.012)
GP FEs	Yes	Yes	Yes	Yes
Quarter FEs	Yes	Yes	Yes	Yes
Outcome Mean	0.17	0.17	0.17	0.17
Adjusted R2	0.011	0.011	0.011	0.011
Observations	8,475,098	8,475,098	8,475,098	8,475,098

Notes: The unit of observation is the quarterly enrollment change for an LSOA-GP. Sample is period when stars were visible. All specifications control for GP age, age squared, and number of practitioners in the GP practice. Standard errors clustered at the GP level in parentheses.

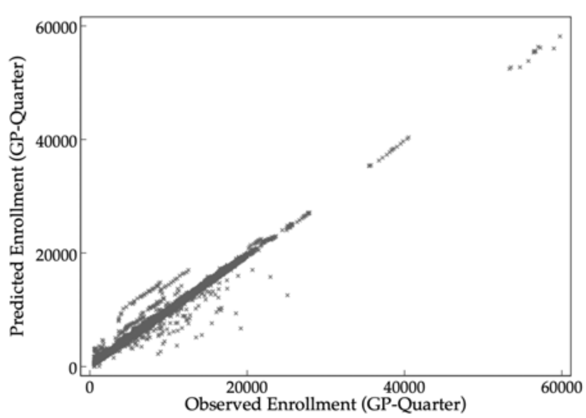
D Additional Structural Model Results

Figure D-1
Smoothness of GP Fixed Effects Around Threshold



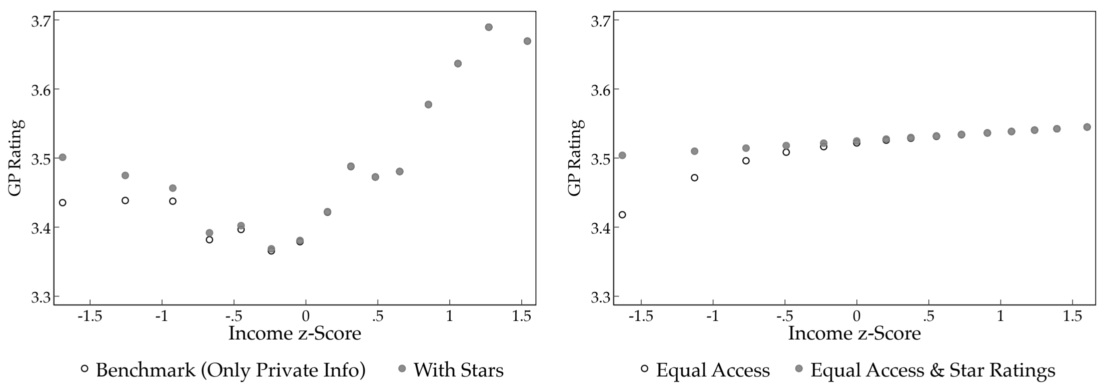
Notes: Dots represent GP fixed effects ξ_j against the distance to the nearest star rating threshold for each GP-quarter. Shaded area is the 95 percent confidence interval from a local polynomial regression using an Epanechnikov kernel on each side of the rounding threshold.

Figure D-2
Model Fit of GP Enrollment



Notes: Figure shows model predictions for enrollment in each GP-Quarter against observed enrollment.

Figure D-3
Relationship between Income and GP Rating
Under Counterfactual Star Ratings and Access to GPs



a. Star Ratings

b. Star Ratings and Access

Notes: Benchmark refers to no star rating system (only private information). Panel (a) shows counterfactual in which all individuals observe the star rating. Panel (b) shows counterfactuals with and without the star ratings when all individuals have the same choice set drawn randomly from the data. All counterfactual simulations examine the long-run (without inertia) averaged over the sample period.

Table D-1
Model Fit with Respect to Targeted Moments in Empirical Model

	Model	Data
$M^1(\Phi)$		
$\tau_{jump,high}$	0.0397	0.0397
$\tau_{slope,high}$	0.5898	0.5898
$\tau_{jump,low}$	0.1504	0.1504
$\tau_{slope,low}$	0.1102	0.1102
$M^3(\Phi)$		
log(Distance)	-0.3888	-0.3889
log(Distance) $\cdot I_\ell$	0.0499	0.0500
r_{jt}	3.2062	3.2060
$r_{jt} \cdot I_\ell$	0.5892	0.5891
m_{jt}	3.2049	3.2057
$m_{jt} \cdot I_\ell$	0.5764	0.5764
Mean GP Experience	-0.0903	-0.0903
Practitioners per 1000 Patients	0.7564	0.7564
$M^4(\Phi)$		
$I^{(4)} = 1$	0.0227	0.0227
$I^{(4)} = 2$	0.0246	0.0246
$I^{(4)} = 3$	0.0257	0.0256
$I^{(4)} = 4$	0.0262	0.0262

Notes: Table shows implied moments based on estimated parameter values from Table 3 and the associated values from the data.

Table D-2
Income-Quality Correlation with Endogenous Capacity Adjustment

Counterfactual	Fixed Capacity	Adjusted Capacity
Benchmark	0.091	0.092
Equal Information	0.069	0.071
Stars	0.070	0.072

Notes: Benchmark refers to no star rating system and baseline choice sets for each individual. Equal information counterfactual gives all individuals the average private signal precision of the top 5% of the income distribution. The stars counterfactual allows individuals to incorporate the information in star ratings.

E Standard Errors for Empirical Model

Standard errors are calculated using the variance-covariance matrix of the GMM estimator

$$\hat{V}_\theta = (D'WD)^{-1}(D'WSWD)(D'WD)^{-1}.$$

Here D is the Jacobian of the vector of moments $M(\Phi)$ with respect to the model parameters (evaluated at the model estimates $\hat{\Phi}$), W is a positive definite weighting matrix, and

$$S = \frac{1}{N}\sum_{i=1}^{N}\left(g_i(\hat{\Phi}) - M(\hat{\Phi})\right)\left(g_i(\hat{\Phi}) - M(\hat{\Phi})\right)'.$$

For each observation, $g_i(\hat{\Phi})$ is the moment contribution evaluated at $\hat{\Phi}$ where

$$M(\hat{\Phi}) = \frac{1}{N}\sum_{i=1}^{N} g_i(\hat{\Phi}).$$

Calculating these observation-level moment contributions is straightforward for all of our moments other than $M^1(\hat{\Phi})$, the difference between the auxiliary model estimates under simulated choices following $\hat{\Phi}$ and the auxiliary model estimates in the observed data.

For these moments, we use the influence function of the OLS estimator to calculate the observation-level moment contributions. Specifically, the auxiliary model estimated is a linear model $y_i = x_i\beta + u_i$, where x_i is a k element row vector that is the i-th row of the design matrix X and y_i is the i-th element of the $N \times 1$ vector y.

The OLS estimator $\hat{\beta} = (X'X)^{-1} X'y$ can be equivalently written using the influence function of the estimator

$$\psi^u = \left(\frac{1}{N}X'X\right)^{-1}(\mathrm{Diag}(\hat{u})\, X)',$$

where $\mathrm{Diag}(\hat{u})$ is an $N \times N$ matrix with the residuals $\hat{u} = y - X\hat{\beta}$ as the diagonal elements and zeros for all non-diagonal elements. Replacing the residuals with predicted values $\hat{y} = X\hat{\beta}$ then gives

$$\psi^y = \left(\frac{1}{N}X'X\right)^{-1}(\mathrm{Diag}(\hat{y})\, X)'$$

where

$$\hat{\beta} = \frac{1}{N}\sum_{i=1}^{N}\psi_i^y.$$