Information and Disparities in Health Care Quality: Evidence from GP Choice in England*

Zach Y. Brown
University of Michigan

Christopher Hansman
Imperial College London

Jordan Keener
University of Michigan

Andre Veiga
Imperial College London

April 2022

[Preliminary and Incomplete]

Abstract

Low-income patients tend to receive lower quality health care. They have limited access to high quality options and—even conditional on access—are less likely to choose high performing providers. We show that differential information about 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 (RD) estimates show that, on average, patients respond sharply to the information on the website, and that this response is almost entirely driven by residents of low-income neighborhoods. We incorporate these RD moments in a structural demand model that allows for consumer inertia as well as heterogeneity by income in baseline information and preferences. Our results indicate that a meaningful fraction of the income-quality gradient could be eliminated by removing informational differences. This suggests that relatively low-cost information interventions have the potential to reduce health care inequalities.

*We thank Radoslaw Kowalski for sharing data. We are grateful to Martin Gaynor, Ben Handel, Carol Propper, and seminar participants at Berkeley, the CEPR virtual IO seminar, Michigan, and Tsinghua for valuable comments.

1 Introduction

There are persistent within-country inequalities in health care throughout the world (Cookson et al. 2021; Hart 1971). Even in high income 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] A key question for policymakers is whether these disparities are solely driven by access to services, or whether they reflect differences in information about health care providers.

In this paper we highlight one example of health care inequality— a positive relationship between income and general practitioner (GP) quality in the United Kingdom—and show that differential information about provider quality is a key cause of the disparity. Our empirical strategy focuses on the temporary presence of a website intended to provide easily accessible information about provider quality. The website, *NHS Choices*, publicly displayed quality ratings of one to five stars for GP practices. We implement a regression discontinuity approach that compares otherwise similar GPs that have a higher or lower number of stars due to marginal differences in the underlying index used to construct the quality ratings.

Our regression discontinuity estimates show that, on average, patients respond sharply to the information provided by the website. While NHS Choices was active, GP practices with higher star ratings experienced greater enrollment growth, even compared to other GPs with nearly identical values of the underlying quality index. Crucially, this increase in demand is effectively entirely driven by residents from low-income neighborhoods,[2] suggesting the presence of meaningful information gaps by income.

To assess the importance of these information gaps in explaining broader health-care inequalities, we incorporate our RD moments into a structural model of GP choice. We allow for inertia and heterogeneity by income in both preferences and the precision of patient information about GP quality. Counterfactual simulations imply that a substantial fraction of the relationship between income and healthcare quality is due to information. These results suggest that relatively low-cost informational interventions can reduce health care disparities.

The core challenge in isolating the role of information comes in separating preferences and geographic or economic constraints from informational barriers. Several aspects of the UK context make it an ideal setting in which to address this problem. First, detailed data

[1] Similar patterns have been found in the United States (AHRQ 2021).

[2] We measure neighborhoods as Lower layer Super Output Areas (LSOAs) throughout.

on GP registration is available, allowing us to track enrollment decisions for all GP offices in England at a fine-grained geographic level. This is combined with a broad range of quality metrics for GPs, based on both patient evaluations and health outcomes. Second, the socialized healthcare system allows us to abstract from the price and insurance-coverage based disparities in quality of care that arise in more market driven healthcare systems like the US. Third, and perhaps most importantly, the unique implementation (and removal) of the star rating system allows us to focus on plausibly exogenous variation in patient information about provider quality.

We begin by providing descriptive evidence on the relationship between neighborhood (LSOA) level income and GP quality. We show that, across a series of metrics, individuals who live in higher-income areas choose higher quality GPs. A portion of this can be explained by access: GPs located in high-income areas tend to be of higher quality. However, this is insufficient to explain the entirety of the relationship. Even conditional on the choice set, those in high-income areas tend to select higher quality GPs. This pattern is consistent with the possibility that high income individuals have more detailed information about provider quality. However, this descriptive evidence is not conclusive. There are a series of alternative explanations that could rationalize patient choices, including unobserved barriers (e.g. discrimination), reverse causality (which might occur, for example, if performing well on various quality metrics is easier when serving a higher income population), and preferences over attributes not captured by observed quality measures.

To isolate the relationship between information and choice—and its role in healthcare disparities by income—we next turn to our regression discontinuity approach. Our strategy is based upon a rating system provided to the public via a website called *NHS choices*. This website allowed patients to directly review GPs, dentists, hospitals, and other providers by leaving written feedback and a score of one to five stars. Between 2013 and the end of 2019, the website provided a summary star rating for each provider, measured as the rolling average score over the past two years, rounded to the nearest half star. We refer to the rolling average as the *index* and the rounded value as the *star rating*. This rounding creates a series of sharp discontinuities. For example, a provider with an index value of 3.76 will appear with 4 stars, while a provider with an index value of 3.74 will appear with 3.5 stars.[3] We compare changes in enrollment patterns for GP practices just above or below these discontinuities.

[3]We are able to observe a panel of summary star ratings directly, and to re-construct the index based on a comprehensive dataset of individual reviews scraped from the website itself.

Our results show that patients respond directly to the information contained within the summary star ratings. While the star ratings were present, providers just above the discontinuities saw significantly larger enrollment growth relative to providers just below. When the summary ratings were removed, in 2019, this difference disappeared. Panel regressions, which focus on practices that experience changes in the summary star rating over time, give the same result. This pattern is consistent with a model in which consumers are uncertain over the quality of GP practices and learn from the information provided by the summary ratings.

Crucially, the changes in enrollment captured by our RD are almost entirely the result of the choices made by patients in low income neighborhoods. Enrollment growth from LSOAs below the median in terms of income is sharply higher for practices just above the discontinuities. There is no observable jump for enrollment growth from LSOAs above the median. In other words, it appears to be low income patients that are uncertain about provider quality, and benefiting from the website. The relationship between enrollment and the continuous index away from the discontinuities further supports this interpretation. Conditional on a star rating, enrollment growth is highly correlated with the underlying index for high income areas, and effectively uncorrelated for low income areas. This suggests that high income patients are aware of (and responsive to) fine differences in quality not captured by the star ratings, while low-income patients are not.

While our reduced form evidence shows that access to information drives consumer choice—and that there are differences in information access across the income distribution—it cannot quantify whether information plays an economically meaningful role in observed health care disparities. As a final step, we construct and estimate a structural model of GP choice. Our model incorporates (and allows for heterogeneity by income in) uncertainty over provider quality. We additionally allow for consumer inertia, to capture stickiness in GP choice over time, and heterogeneity in patient preferences for quality and observable GP characteristics.

We estimate the model using an indirect inference approach (Gourieroux et al. 1993) that incorporates our RD results (see Allende 2022; Fu and Gregory 2019, for recent papers incorporating RD moments into indirect inference approaches). We consider counterfactual simulations in which informational differences across income groups are eliminated, as well as in which differential barriers to access are removed. We find that equating information can remove a substantial fraction of the relationship between income and GP quality, on the order of 8%. Furthermore, the entire relationship disappears once accounting for information and access. This suggests that differences in preferences are not an important driver of the observed income-quality gradient. Taken as a whole. our results indicate that relatively low cost informational interventions may be effective in weakening health care inequalities, but that access remains a key barrier.

Our primary contribution comes in providing evidence on—as well as modeling and quantifying the role of—informational differences as a driver of health care inequalities. 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), including in 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 also finds evidence that information about health plan quality affects choices (e.g. Jin and Sorensen 2006; Dafny and Dranove 2008; Werner et al. 2012; Kling et al. 2012). Our innovation comes in highlighting information as an important source of disparities in health care.

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 Dynarski and Scott-Clayton 2006; Dynarski et al. 2021; Bettinger et al. 2012; Hastings et al. 2015; Oreopoulos and Ford 2019). We document this channel in the health care context, where inequality is a first order concern and there is substantial evidence of significant 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), Lewis and Zervas (2016), Luca (2016), Luca and Zervas (2016), Reimers and Waldfogel (2021), Anderson and Magruder (2012) and more, particularly those that explicitly incorporate learning models (e.g. Newberry and Zhou 2019). Our primary contribution is to highlight the importance (and promise) of such systems in driving demand for healthcare markets and in using a review-based identification 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 practice choice in the presence of uncertainty and a rating website. Section 4 introduces and present results from our regression discontinuity design. Section 5 presents our structural model and shows counterfactual simulations. Section 6 concludes.

2 Background, Data and Disparities in GP Quality

In this section we begin by providing background on the process of choosing a general practitioner (GP) in England, include details on the website at the heart of our analysis. We then outline the data used in our analyses and provide descriptive evidence on the existence of disparities in GP Practice quality by income.

2.1 Background on GP Practices and the NHS Choice Website

Choosing a GP Practice

In England, as in many other countries, General Practitioners (GPs) act as the first point of contact for patients experiencing health issues or seeking routine preventative care. GPs provide a wide range of services, including checkups, screenings, vaccinations, simple surgeries, and referrals to specialists and hospitals. They are typically organized into practices comprising several physicians, which contract directly with the National Health Service (NHS).

All individuals in the UK can register with a GP practice free of charge, and the right to choose a practice is directly outlined in the constitution of the NHS.[4] Absent capacity constraints or a set of specialized circumstances, practices must generally accept patients who register.[5] The process of registration is straightforward and can often be completed online.

4. 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." See https://www.gov.uk/government/publications/the-nhs-constitution-for-england.

5 https://www.gov.uk/government/publications/the-nhs-choice-framework provides the set of circumstances in which a GP Practice may refuse a registration or limit choice of nurses and doctors. As of January 2015, patients can choose a practice outside their designated geographic area. However 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.

Patients are free to switch GP practices at any time, and medical records are automatically transferred.

Online Reviews: NHS Choices

In an effort to facilitate informed decision making, the NHS runs a website (www.nhs.uk) that provides details on healthcare and pharmaceutical services alongside general purpose medical information. A key component of the website is a guide to GP practices, hospitals and other healthcare providers. Through the website, patients are able to access their records, make appointments, order repeat prescriptions, and, importantly for our purposes, search for and begin the registration process for a GP practice.

The core focus of our analysis is on a rating system for GP practices and other types of providers that is included on the website, and was initially referred to as *NHS Choices*. This system allows patients to leave a written review of a provider and to provide a ranking from one to five stars.[6] From prior to 2015 to the end of 2019, these rankings were used to construct *summary star ratings* which were displayed prominently on the provider's page as well as in search results. These summary star ratings were calculated as the average rating over the past two years, rounded to the nearest half-star (e.g. 3.5 stars). An example of the page for a particular GP Practice during this period is presented Panel a of Figure A-1. The summary star rating can be seen in the upper right-hand corner.

At the end of 2019, the website re-branded from "NHS Choices" to the "NHS Website" and the website removed the summary star ratings from the provider pages (see Figure A-1 Panel b). While individuals could still theoretically find individual rankings and manually calculate the average, the information was significantly more difficult to access. Our primary analysis considers the period prior to January 2020 in which the star ratings were directly displayed, but we also consider the period after the ratings were removed as a falsification exercise.

[6] In addition, patients can leave reviews through other websites, which are then incorporated into the NHS choices website. These reviews are also included in our analysis.

2.2 Data

Review and Star Ratings Data

We obtained individual reviews and rankings for each GP practice from the NHS Choices website for the period April 2016 to December 2021. Since reviews older than two years are removed from the website (and indeed deleted by the NHS), we combine this with separate data that was collected for the period May 2013 to April 2016.[7] The combined dataset contains all reviews from May 2013 to December 2021. We use these to construct a monthly panel from May 2015-December 2021 that contains two key variables for each practice j: the index r_{jt}, which is the precise average across all rankings (based on a two year moving average) and the rounded star ranking s_{jt}, which is simply r_{jt} rounded to the nearest half star.

There are two main concerns in interpreting an index based in user feedback. The first is whether it captures a meaningful measure of healthcare and clinical quality. Analysis of written reviews indicates that patients judge providers based on a wide variety of concerns, including quality of the medical services provider, amount of bureaucratic red tape, bed-side manner, and the quality of the facilities (Kowalski 2017). This suggests that at least a portion of reviews speak to clinical practices. We verify this by examining the correlation between the index r_{jt} and other subjective and objective measures of quality in Appendix Table A-1. The relationship with various measures from survey data is high: we find a correlation of 0.52 with a measure of overall experience based in representative patient surveys. r_{jt} is also highly correlated with the NHS Quality and Outcomes Framework (QOF), a measure of provider quality that includes objective clinical quality indicators. All correlation coefficients are statistically significant, providing evidence that patient feedback reflects a deeper notion of quality.[8]

The second concern is that there may be credibility issues, particularly if fake ratings

[7] We thank Radoslaw Kowalski for providing these data. See Kowalski (2017).

[8] The GP Patient Survey is an independently-run representative annual survey of over 1 million individuals that is run on behalf of NHS England.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. The Quality and Outcomes Framework (QOF) is a system commonly used for performance pay of GPs. Most GP funding is based on the the number of patients enrolled at the practice, weighted by patient complexity. This is supplemented with performance pay based on the QOF if GPs voluntary participate. 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 are relatively difficult to compare across GP practices.

are common as has been documented in other review systems (Mayzlin et al. 2014; Luca and Zervas 2016). However, unlike many online feedback mechanisms, provider ratings are government sanctioned and moderated by the NHS, which collects information on each individual leaving a review, including email and IP addresses.[9] This moderation is likely to ensure that reviews are informative, and discourage explicit gaming by providers or their employees. The fact that ratings are highly correlated with independent measures of quality further helps assuage credibility concerns. In addition, because our primary empirical strategy is based in a regression discontinuity, the main threat for our purposes is the presence of differential fake ratings near rounding thresholds for summary star ratings. We consider this possibility in more detail when discussing the plausibility of our approach in Section 4

The star rating for GPs with very few reviews carries limited information. Individuals are likely aware of this given that the number of reviews is prominently displayed near the star rating on the website. This can also be seen by noting that average ratings are less correlated with other measures of quality when the number of reviews is small (see Appendix Table A-1). As a result, our preferred specifications for our reduced form analyses consider only GP Practice - Months with at least 5 reviews (6,673 practices satisfy this restriction at some point). We include robustness exercises that relax this restriction.

GP Enrollment and Characteristics Data

We match our review and star ratings data with enrollment data for the universe of GP practices in England. These data are extracted from the Primary Care Registration database within the National Health Application and Infrastructure Services (NHAIS) system. For each GP practice, we observe quarterly enrollment by age, gender, and Lower Level Super Outut Area (LSOA). LSOAs are fine-grained geographic areas with an average of about 700 households (2,000 individuals), and are roughly analogous to census block groups in the US. We use this to construct a quarterly panel of adult enrollment at the GP practice$\times$LSOA level. We merge on time varying LSOA characteristics, including income, health, education, and employment, as well as quarterly data on GP practice characteristics such as the number of practitioners, mean experience of practitioners, and the age of the practice. In addition, we geocode addresses for all GP practices, allowing us to calculate the distance to the centroid of each LSOA.

GP enrollment is highly persistent from quarter to quarter, as patients remain enrolled

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

with their current GP in the absence of any medical care (and may not switch unless they move, have a new health issue, or are dissatisfied with their current GP). As a result, we construct and focus on the quarterly change in enrollment at the LSOA level, as well as enrollment growth in percentage terms. 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.[10]

2.3 Summary Statistics

We show summary statistics on GP practice enrollment, patient demographics, and reviews in Table 1. The first two columns show data from the period in which summary star ratings were visible. Our unrestricted sample includes 7,635 unique GP practices, over 18 million GP×LSOA×quarter observations and over 350,000 individual star reviews. About 8,000 patients were enrolled at the average GP and the average quarterly enrollment change from an LSOA is just below 0.2 patients, or 2.4 percent. Average enrollment was slightly higher in the post-star rating period, at just over 9000. 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.[11] Appendix Figure A-2 shows a histogram of distance between LSOA centroids and chosen GP practices. The median individual lives only 1.4 km from their GP, implying distance is an important determinant of choice.

The distribution of individual rankings and the average index value r_{jt} at the GP-month level are shown in Figure A-3 and Figure A-4, respectively. Most individual rankings are either 1 star or 5 star, however Figure A-4 indicates that the distribution of the index across GP practices is relatively smooth, including near the star thresholds. The mean GP practice has an index of 3.2 stars with a large degree of dispersion—the standard deviation across practices is 1.0 stars.

[10] Although GP mergers are thought to be common, there is no explicit way to identify them in our data.

[11] The 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.

Table 1
Summary of GP Enrollment and Characteristics

	Period with Star Ratings		Period without Star Ratings	
	Mean	SD	Mean	SD
GP Enrollment:				
Total Enrollment (100s)	80.75	50.92	91.38	60.50
LSOA Enrollment	0.58	1.61	0.54	1.61
Quarterly LSOA Enrollment Change	0.17	2.08	0.09	1.81
Quarterly LSOA Enrollment Growth	2.36	26.25	2.55	25.78
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
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
Unique GPs	7,635			
Total GP Observations	18,415,832			
Individual Reviews	356,983			

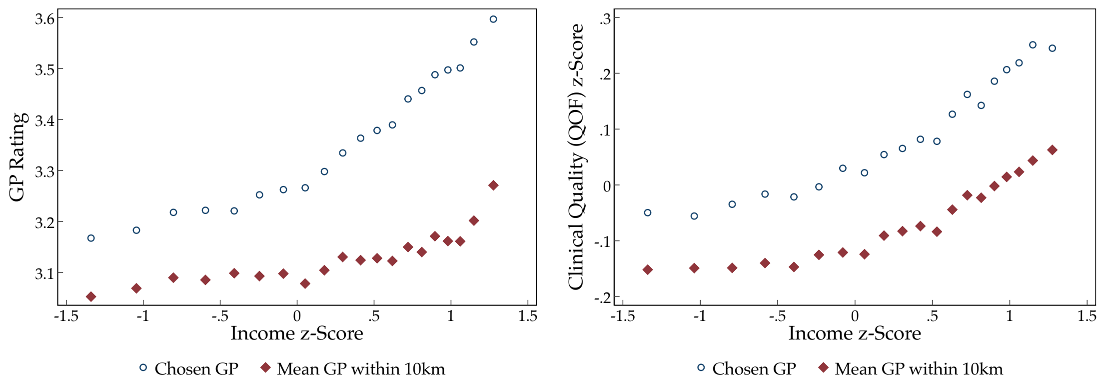
Notes: Sample includes GPs in England from May 2015 to December 2021.

2.4 Descriptive Evidence on Disparities in GP Quality By Income

We begin our analysis by showing 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 a measure of LSOA level income—the normalized income deprivation index—and the star rating index r_{jt}. There is a notable upward trend. Individuals living in ares that are one standard deviation above the mean in terms of our income measure attend GP practices that are rated roughly 0.3 stars higher relative to those living in LSOAs that are one standard deviation below the mean. This pattern is not an artifact of the star rating system. Panel b of Figure 1 shows a shows the relationship between income and the Quality and Outcomes Framework clinical score, a more objective measure of GP quality. The relationship between income and GP quality is similar, indicating the existence of a meaningful underlying disparity in quality by income. This is largely consistent with other work documenting inequality in health care in the UK (e.g. O'Dowd 2020; Scobie and

Morris 2020).

Figure 1
Relationship between Income and GP Quality



a. Average patient review
b. Quality and Outcomes Framework Score

Notes: Binscatter plots show relationship between the z-score of patient LSOA income and the GP quality of the chosen option and the GP with median quality within 10km. In Panel a, quality is measures as the average patient rating on the NHS website. In Panel b, quality is measures using the Quality Outcome Framework (QOF) clinical quality score. Results are weighted by the quarterly change in enrollment.

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 our LSOA level income measure and a proxy for the quality of the choice set facing the individuals in the LSOA: the mean of the GP rating index for all GPs within 10 kilometers of the LSOA centroid. There is again a noticeable upward trend: high income areas tend to have higher quality GP practices nearby. However, the key point is that the relationship between chosen quality and income is significantly steeper than that between choice-set quality and income. The blue dots rise faster than the red as LSOA level income increases. In other words, even conditional on access, high income individuals appear to choose higher quality GP practices.

While this descriptive evidence suggests that choice plays a role in explaining GP quality disparities, it does not necessarily indicate that information differences are a key factor. These observed disparities could be the result of differences in preferences, differences in access that are unobservable (for example, discrimination or registration processes that are favorable to high income individuals) or other factors. In Section 4, we provide regression discontinuity based evidence that information is indeed an important component. To decompose and quantify the determinants of the income-quality gradient, we estimate an information based demand model that accounts for differences in the choice set between low-income and high-income patients in Section 3.

3 A Model of Demand For GP Practices with Learning

Our central questions are whether (i) patients, in general, lack relevant information on quality when choosing a GP and (ii) these information gaps are larger for low income individuals. The challenge comes in separating information from other barriers or confounds that lead to heterogeneity in choice. For any given quality metric, the simple fact that some patients choose lower ranked providers is not necessarily indicative of an information gap. These choices could reflect heterogeneity in preferences, or potentially unobservable differences in costs, access, or constraints.

In this section, we present a simple model to show that the presence of information gaps—which we define as imprecision over provider quality—has sharp implications for patient choice in the presence of a star rating website like *NHS* choices. This simple model shows how information and preferences can be separately identified, motivating our regression discontinuity analysis in section 4.1 and our empirical model in Section 5.

The key insight is that patients will only respond sharply to star ratings if they have imprecise information about quality. With perfectly privately informed patients, we should not expect discontinuous jumps in enrollment for practices with ratings that are rounded up versus rounded down (e.g. $r_{jt} = 3.76$ rounded to $s_{jt} = 4$ versus $r_{jt} = 3.74$ rounded to $s_{jt} = 3.5$). On the other hand, a Bayesian patient with imprecise private information will incorporate information from star ratings and have discretely higher beliefs about the quality of practices with better star ratings. In addition, our model implies that heterogeneity in the precision of beliefs translates into heterogeneity in responses to star ratings. If there are disparities in the precision of information about provider quality, we should expect these to lead to differences in enrollment across star ratings.

3.1 Patient Beliefs About Practice Quality

Let true quality of GP $j \in \mathcal{J}$ be given by r_j. Suppose the rounded star rating s_j is public information. Absent any private information, all agents have prior

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

Now suppose each individual receives a noisy private signal about the quality of provider j, which could be the result of their own research, insight from social networks, past experience, or any other channel. We model this as

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

with $\epsilon_{ij} \sim N(0, \sigma_i^2)$. Given the prior and the signal, individuals form posterior beliefs using Bayes' rule. Expected quality is therefore given by

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

The weight on the private information, α_{ij}, is given by

$$\alpha_{ij} = \frac{\sigma_s^2}{\sigma_{ij}^2 + \sigma_s^2}. \tag{2}$$

This weight captures the simple intuition behind our model. Individuals with perfectly precise private signals of quality ($\sigma_i^2 = 0$) place no weight on the information contained in the star rating. Alternatively, individuals with imprecise private signals (large values of σ_i^2) place significant weight on star ratings.

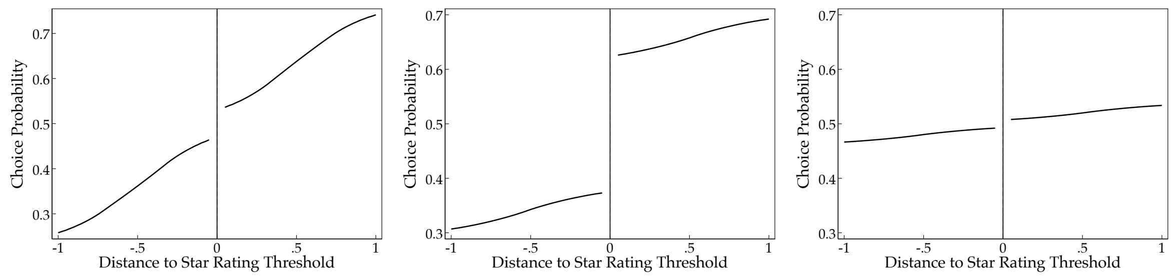
3.2 Patient Utility

We assume that patients are risk neutral, value quality r_j, and have a standard normal individual taste shocks ν_{ij}. Given these assumptions, expected utility can be expressed as:

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

where the joint error, defined as $\nu'_{ij} \equiv \beta\alpha_{ji}\epsilon_{ij} + \nu_{ij}$, is normally distributed with mean zero and variance $\beta^2\alpha_{ji}^2\sigma_{ij}^2 + 1$. Choice probabilities therefore follow a multinomial probit model.

Figure 2
Simulated Response to Star Rating Threshold



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

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

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

Notes: Charts show simulated outcomes for individuals choosing between two GPs with $r_j \sim N(0,1)$. Simulation assumes stars are $s_j \in \{0,1\}$, e.g. can be high or low depending on whether $r_j > 0$. In addition, we assume $\hat{\sigma}_j^2(N_j) = 0$ for simplicity.

3.3 Implications for Ratings and Patient Choice

This simple model has direct predictions about the relationship between GP choice, true quality r_j, and rounded ratings s_j, given a patient's preference for quality and the precision of their information (parameterized by β and σ_i, respectively). These predictions are summarized in the simulation results shown in Figure 2, which show the relationship between demand (choice probabilities) and the underlying quality r_j. The simulations are based on a version of the model with two providers and two possible rounded star ratings: $s_j \in \{0,1\}$.[12] We simulate many individuals, each with a different choice set, and average over the simulations.

For patients with relatively precise private information about quality—for example, high income individuals that can easily gather information from their social networks—there is a strong relationship between underlying provider quality r_j and demand. This is visible from the steep slope of the black lines in panel (a) of Figure 2. Furthermore, given the quality of their signal, they do not significantly adjust their beliefs on the basis of star star rating itself. This is reflected in the relatively modest jump at the star rating threshold.

In contrast, individuals with less precise private signals, for example, low-income in-

[12]For these simulations, r_j 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.

dividuals that do not have access to other sources of information about quality, show a relatively flat slope in the relationship between choice and r_j away from the threshold, and a more striking 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 key takeaway is that the combination of the slope (between demand and r_j) and the jump at the threshold allows us to disentangle patients preferences for quality from the imprecision of their information. In other words, the combination of patients' responses to underlying quality r_j and the star ratings s_j. 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

4 Reduced Form Evidence on Patient Responses to Star Ratings

In this section, we provide reduced form evidence that (i) patients lack relevant information on quality when choosing a GP and (ii) that these information gaps are larger for low income individuals. We begin by using a regression discontinuity approach and supplement this with a panel regression strategy focusing on GP practices that experience changes in rounded star ratings over time.

4.1 RD Methodology

Our goal is to recover the causal effect of star ratings on GP practice enrollment in general, and to test whether this effect differs for those living in high versus low income areas. In the context of our model, this translates to testing whether patients are fully informed about provider quality, and whether there are differences in the precision of information by income.

We implement a relatively standard regression discontinuity approach, exploiting the fact that star ratings (s_{jt}) on the NHS Choices website are rounded to the nearest half star. This means that two GP practices may display different star ratings even if the average ratings are very similar. We consider the underlying index value r_{jt} as a running variable, and compare enrollment just above versus just below each rounding threshold (which exist at every half star interval between 1.25 and 4.75).

Appendix Figure A-5 displays the basic patterns our strategy captures, plotting GP practice $\times$ LSOA level changes in enrollment across the rounding thresholds during the period in our sample in which reviews were active. At each threshold, we see a discontinuous jump. Because our focus is not on heterogeneity across the range of star ratings, our main specification collapses the data and jointly estimates the effect of crossing any threshold.

Formally, we let y_{jlt} represent an enrollment outcome at the practice j, LSOA l and quarter t level and $y_{jlt}(s)$ represent the potential outcome with rounded star rating s. Letting c_s represent the rounding threshold just above s (i.e. $c_s = s + 0.25$), our target parameter is

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

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 this, we first stack our data and normalize all thresholds c_s to 0. As a running varaible, we consider $r^0_{jt} = r_{jt} - c_s$, the distance to the relevant threshold.

We take a non-parametric local linear (or local polynomial) approach to estimating this parameter following (Cattaneo et al. 2019). Given a bandwidth h, we estimate separate weighted least squares regressions of y_{jlt} for observations with $r^0_{jt} > 0$ and $r^0_{jt} < 0$, weighting each observation according to some kernel function $K(\frac{r^0_{jt}}{h})$. We recover the intercepts α_+ (using observations with positive values of r^0_{jt}) such that $\hat{y}_{jlt} = \hat{\alpha}_+ + \hat{\beta}_+ r^0_{jt}$, and α_- (using observations with negative values of r^0_{jt}) such that $\hat{y}_{jlt} = \hat{\alpha}_- + \hat{\beta}_- r^0_{jt}$. Our estimate is then

$$\hat{\tau} = \hat{\alpha}_+ - \hat{\alpha}_-.$$

Our baseline approach uses a triangular kernel, although we consider alternatives for robustness. When considering a rectangular kernel, the above simplifies to estimating the following linear regression (for observations with $r^0_{jt} \in [-h, h]$):

$$y_{jlt} = \alpha_- + \tau \mathbb{1}\{r^0_{jt} > 0\} + \beta_- r^0_{jt} + (\beta_+ - \beta_-)\mathbb{1}\{r^0_{jt} > 0\} \times r^0_{jt} + \varepsilon_{jlt}. \tag{5}$$

We select symmetric MSE-optimal bandwidths following Calonico et al. (2014) and Calonico et al. (2019), again considering several alternatives for robustness. We compute standard errors clustered at the practice level using the plug-in residual approach outlined in Calonico et al. (2019). In our baseline specifications, we further include a vector of covariates for the GP practice, X_{jt}, including GP age and number of reviews, as well as cutoff fixed effects,

δ_c.[13] We show that our results are robust to excluding these covariates.

Our key identification assumption is that the relevant average potential outcomes functions are continuous at each threshold. That is, that $E[y_{jlt}(s)|r_{jt}]$ and $E[y_{jlt}(s+0.5)|r_{jt}]$ are continuous at the point $r_{jt} = c_s$ for each star rating s. This might fail if there is endogenous sorting of GP practices across thresholds.[14] Endogenous sorting could result from GPs that strategically respond to star ratings threshold. For instance, GPs right below the threshold may attempt to provide better service to patients in order to increase their star ratings. Alternatively, some more fundamental feature of the rating system, for example, the number of reviews given to different types of providers, or a tendency of certain GP practices to manipulate reviews, could generate discontinuities.

Our analysis proceeds in three steps. We first provide evidence for the plausibility of our identifying assumption. We then show that the 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

We show three pieces of evidence in support of our identifying assumption: (i) manipulation tests in the spirit of McCrary (2008) to evaluate potential jumps in the distribution of r_{jt} across the threshold, (ii) tests of the smoothness of covariates across the threshold, to consider the existence of discrete jumps in the observable features of GP practices above vs. below each threshold, and finally (iii) tests of differences in the distribution of reviews before vs after the review system was visible, to test for manipulation of the review system.

Manipulation Tests

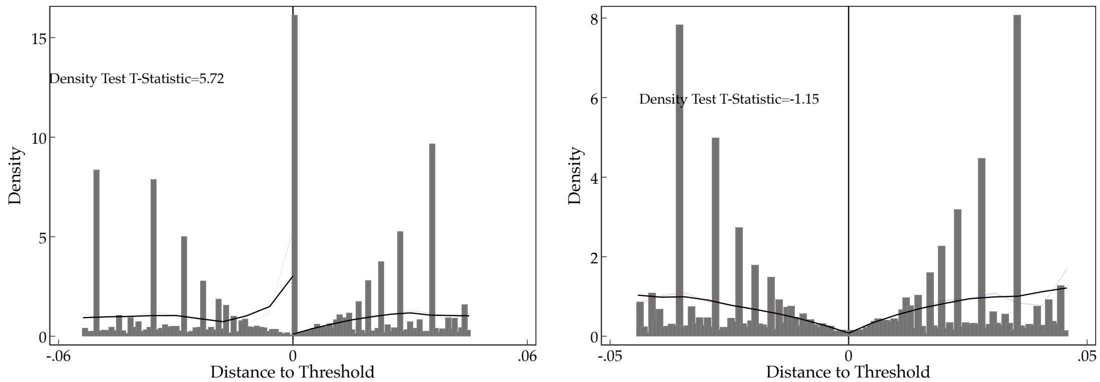
We implement the tests outlined in Cattaneo et al. (2018) based on local polynomial density estimators. For these tests we use the suggested MSE optimal bandwidth and show unrestricted robust bias-corrected t-statistics. We show two versions of plots and t-statistics for these tests in Figure 3.

As panel a shows, there is one generic source of bunching in the distribution of r_{jt}.. Because r_{jt} is calculated as the average of integer rankings by patients, the ratings are likely

[13]The inclusion of threshold fixed effects improves precision by controlling for the fact that the outcome level is different for each threshold.

[14]See, for instance, Urquiola and Verhoogen (2009) and Bajari et al. (2011).

Figure 3
Density Tests of GP Average Reviews Around Star Rating Thresholds



a. Full Sample

b. Estimation Sample

Notes: Histograms and polynomial density estimates for practices above and below star rounding thresholds. Unrestricted robust bias-corrected t-statistics following Cattaneo et al. (2018) shown in upper left hand corner.

to be exactly equal to one of the thresholds with a relatively small number of reviews. Consequently, when using the full sample we reject the null that the density of reviews is continuous across the threshold ($t = -5.72$). To ensure that this feature of the rating system does not drive our results, our primary estimation sample drops all observations with r_{jt} exactly equal to one of the ratings thresholds (although results are effectively the same whether these observations are included or not). The results from this estimation sample are shown in panel b. In this sample we do not find evidence of a discontinuous jump in the density of reviews at the threshold ($t = -1.15$).

Smoothness of Covariates

Appendix Figure A-6 shows that observable characteristics of GP practices are continuous through the thresholds. For out tests, we implement the RD methodology described above at the GP level, with four different types of GP practice characteristics on the left-hand side: (i) the number of months the GP has been active, (ii) a survey based outcome of patient trust, (iii) the QOF clinical score, and (iv) 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. This provides reassurance that there is no discrete change in provider type across the threshold.

Distribution of Patient Ratings

As a last test, we examine whether the distribution of ratings changed after the star ratings were removed from the website in 2020. As seen in Appendix Figure A-7, there was no statistically significant difference in the distribution of reviews before and after the star ratings thresholds were displayed, providing evidence against fake reviews or strategic behavior on the part of GP practices in an effort to game the rounded star ratings.

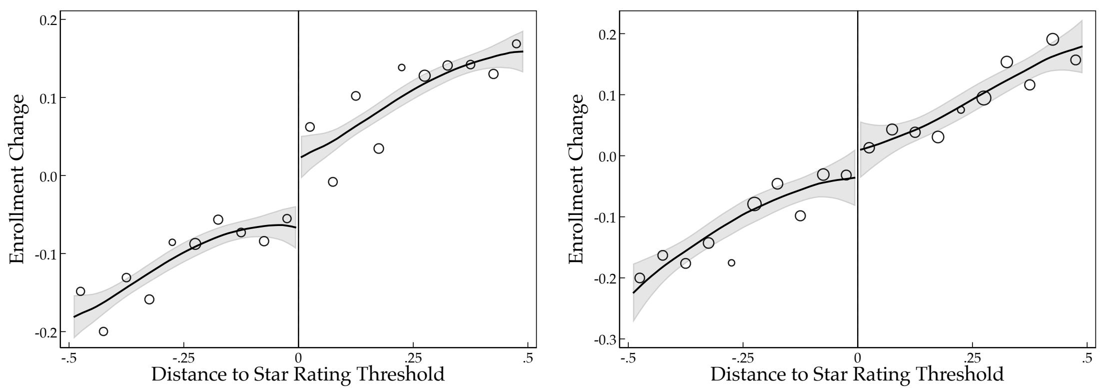
4.3 Regression Discontinuity Results

Average Impact of Star Ratings on Enrollment

The baseline impact on enrollment for our full estimation sample is shown graphically in Figure 4 Panel a. The chart shows the main outcome of interest, enrollment change, as a function of the distance to the star rating threshold. We show both a binned scatter-plot and a local linear smoothing on each side of the discontinuity. The relationship between average reviews and the change in enrollment is roughly linear with a positive slope on both sides of the threshold. This positive slope indicates that patients value—and have some independent information—about the notion of quality captured by the index r_{jt}. There is a clear jump in the outcome at the star rating threshold indicating that a higher star rating leads to a meaningful effect on demand.

The RD results corresponding to this figure are presented in the first two columns of top panel of Table 2. Our preferred estimates imply that a half star jump in ratings increases the quarterly change in enrollment from an LSOA by 0.13 patients. For comparison, the mean LSOA enrollment at a GP practice is 0.58, implying that a half star increases enrollment by 22%. This effect is statistically significant at the 5 percent level. Robust, bias corrected 95 percent confidence intervals do not contain 0. Column 2 shows that our results are not solely the result of our bandwidth choice. Our estimates are similar when using the bandwidth selection procedure of Imbens and Kalyanaraman (2012). These results suggest that, on average, patients are not perfectly informed about the quality of GP practices, and that they value and respond to the information provided by the rounded star-rating system.

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



a. Visible star ratings

b. No star ratings

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

Falsification Tests: Enrollment After the Website Removed Star Ratings

In January of 2020, rounded star ratings were removed from the website. This provides us with an ideal falsification test. If the enrollment effects we estimate are truly driven by rounded star ratings, we should expect them to disappear after the rounded star ratings are removed. On the other hand, if these effects are spuriously driven by some discrete change in practice type at the threshold, the effect would not disappear.

Panel b of Figure 4 shows that the results indeed disappeared after the ratings were removed from the website. While the relationship between reviews and GP demand is similar on either side of the threshold, there is no visable discontinuity when using data after January 2020. Consistent with this graphical evidence, our RD estimates (shown in columns 3 and 4 of Table 2) indicate that there is no statistically significant effect in this period. This provides further evidence that our estimates are indeed driven by the impact of the star ratings themselves.

Differences by Income

We now turn to the central focus of our paper, heterogeneity in the impact of information by income. Our basic approach is to separately implement our RD strategy on subsamples

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

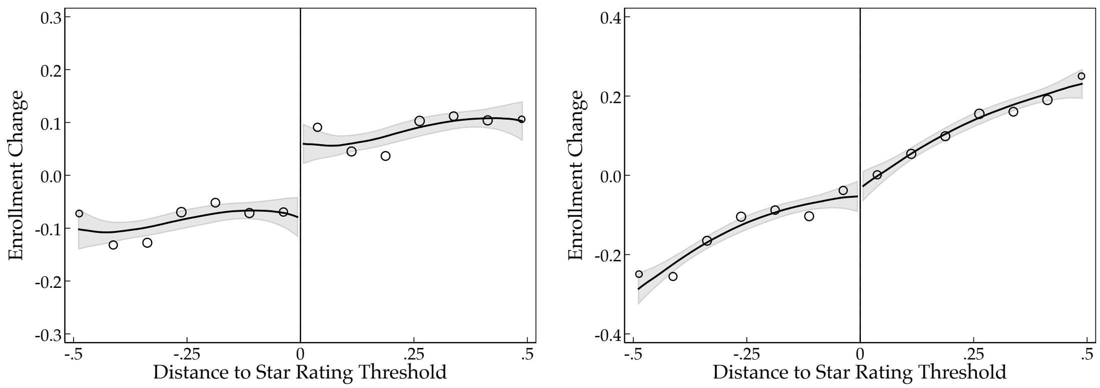
Notes: Dependent variable is change in GP enrollment. Sample excludes quarterly GP observations with fewer than five reviews. Each regression controls for average review score using a local linear regression with triangular kernal. Optimal bandwidths are calculated for each specification. Bottom panel uses CCT bandwidth. Low (high) income is defined as those in an LSOA with below (above) median income. Standard errors are in parentheses and clustered by GP.

of high versus low income LSOAs.[15]

Figure 5 presents graphical evidence of the key result: the impacts of the star rating website are largely driven by low income LSOAs. Panel a shows clear evidence of a discontinuity in enrollment changes for low income LSOAs, defined as those below the median according to our income measure. In contrast, there is no evidence of an effect in Panel b which shows above median LSOAs. The regression results in panel b of Table 2 confirm this results. For individuals in neighborhoods with below median income, the effect of crossing the star ratings threshold on enrollment change is 0.19 and is significant at the 1 percent level. For the individuals in neighborhoods with above median income, the effect is not statistically significant. Furthermore, neither result is statistically significant in the period without star

[15]Our basic measure of income is LSOA level Income Deprivation, which measures the fraction of very low income individuals in an LSOA.

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



a. Low Income

b. High Income

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

ratings, providing further evidence for the validity of our strategy.

While not the focus of our RD Approach, the relationship between demand and average ratings away from the threshold provide insight into agents preferences for high quality GP practices. For high income LSOAs, we see relatively steep slopes away from the thresholds, despite the lack of a discontinuity at the threshold. Alternatively, we see relatively flat slopes for low income LSOAs. In the context of our model, this suggests that both low and high income individuals value quality, but that high income individuals are able to collect more precise information from sources other than the rating website.

In Appendix Figure A-8 we show evidence that similar patterns hold across other measures of socioeconomic status. We find that individuals with below median education, employment, or health all respond to the star ratings. In contrast, individuals with above median education, employment, or health do not have a statistically significant response at the threshold, but do demand quality away from the threshold. We also show more fine-grained evidence on the income distribution in Figure A-9, which shows separate plots quartile by quartile.

Robustness for RD Approach

In Appendix Tables A-2 and A-3 we show that our RD approach is not sensitive 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 alternative specifications, we observe the same patterns. Higher star ratings generate increased enrollment, and this enrollment is largely driven by residents of low-income LSOAs.

4.4 Alternative Identification Strategy: Panel Regressions

As a final piece of reduced form evidence, we show that enrollment responds to *within*-firm 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 effectively constant). 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:

$$y_{jlt} = \gamma s_{jt} + X'_{jt}\beta + \alpha_j + \delta_t + \epsilon_{jlt} \tag{6}$$

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.[16] We also control for time-varying characteristics of the GP, X_{jt}, including age of the GP, 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 RD approach, 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 A-4, are consistent with our RD strategy. In the full sample, star ratings have a positive and significant effect the quarterly change in enrollment. In Column 2 of Appendix Table A-4 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.

[16] 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.

Overall, 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 approach. Of course, 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. Still, the consistency across approaches is reassuring, and the fact that patient demand is responsive to within-practice changes in ratings further confirms that our RD approach is not driven by some spurious cross-sectional sorting around the threshold. However, given the stronger assumptions, we focus primarily on incorporating our RD approach into the structural model presented in the following section.

5 Empirical Model of Demand

We now present an empirical framework that builds on the model of patient information and demand presented Section 3.1, while also seeking to capture other important determinants of demand in the context of GP choice. This includes accounting for the presence of consumer inertia in health provider choice (see, e.g. Raval and Rosenbaum 2018; Shepard 2022).[17]

If an individual makes an active choice, we assume utility follows the specification outlined in Section 3.1 with a few slight generalizations. Specifically, individual i in LSOA ℓ and quarter t has expected utility from choosing GP $j \in \mathcal{J}_{\ell t}$: given by

$$\mathbb{E}[u_{i\ell jt}] = \beta_{1\ell}\left[\alpha_{\ell t} r_j + (1-\alpha_{\ell t})\mathbb{E}[r_j|s_{jt}]\right] + f(d_{\ell j}, X^d_{\ell t}; \beta_2) + X'_{jt}\beta_3 + \xi_j + \nu_{i\ell jt}.$$

Here $f(d_{\ell j}, X^d_{\ell t}; \beta_2)$ is a function of distance $d_{\ell j}$ from LSOA ℓ and j interacted with LSOA-level income, age and health. X_{jt} is a vector of time varying GP characteristics that includes mean physician age within the practice and the number of practitioners per patient. ξ_j are GP fixed effects capturing measures of quality that are observed by the individual but not the researcher. We let preference for quality vary with LSOA level income $\beta_{1\ell} = \beta_{10} + \beta_{11} I_{\ell t}$.

$\nu_{i\ell jt}$ is a composite error that incorporates both individual i's taste shock and their private

[17] A literature has also explored the implications of consumer inertia in other settings including health insurance choice (e.g. Handel 2013; Ho et al. 2017), electricity (e.g. Hortaçsu et al. 2017), and financial products (e.g. Luco 2019).

signal about the quality of GP j. Following the theoretical model, the variance of this error is given by $Var[\nu_{i\ell jt}] = \sigma^2_{\nu_\ell} = \beta^2_{1\ell}\alpha^2_{\ell t}\sigma^2_{ij} + 1$. For computational tractability, we assume that this error follows an EV1 distribution. We parameterize the variance of each individual's private signal as a function of LSOA level income:

$$\sigma_{ij} = \exp[\gamma_0 + \gamma_1 I_{\ell t}]. \tag{7}$$

γ_0 and γ_1 determine the precision of an individuals private information about quality and how that precision changes with income. Therefore, the weight that patients put on their prior, $\alpha_{\ell t}$, is given by

$$\alpha_{\ell t} = \frac{\sigma^2_s}{\exp[\gamma_0 + \gamma_1 I_{lt}]^2 + \sigma^2_s}. \tag{8}$$

The variance of the information contained in star ratings, σ^2_s, depends on the variance of quality within star rating bins and is calculated directly from the data.

Only a modest fraction of individuals switch GP practices in any given quarter. We model this as an exogenous re-optimization probability that depends on observable characteristics of the LSOA $\varphi_{\ell t}$. This can be thought of as jointly capturing the probability of moving, receiving a health shock, and any other factor that causes individuals to consider selecting a new GP practice. Specifically, we set

$$\varphi_{\ell t} = \frac{\exp[X^{a}_{\ell t}{}'\theta]}{(1 + \exp[X^{a}_{\ell t}{}'\theta])}$$

Given these specifications, we may write the market share for provider j within LSOA l in quarter t as[18]:

$$S_{\ell jt} = \varphi_{\ell t}\frac{\exp\left[\frac{1}{\tilde{\sigma}_{\nu_\ell}(\beta_{1\ell},\gamma)}\left(\beta_{1\ell}\left[\alpha_{\ell t}r_j + (1-\alpha_{\ell t})\mathbb{E}[r_j|s_{jt}]\right] + f(d_{\ell j}, X^d_{\ell t};\beta_2) + X'_{jt}\beta_3 + \xi_j\right)\right]}{\sum_{k\in\mathcal{J}_{lt}}\exp\left[\frac{1}{\tilde{\sigma}_{\nu_\ell}(\beta_{1\ell},\gamma)}\left(\beta_{1\ell}\left[\alpha_{\ell t}r_k + (1-\alpha_{\ell t})\mathbb{E}[r_k|s_{kt}]\right] + f(d_{\ell k}, X^d_{\ell t};\beta_2) + X'_{kt}\beta_3 + \xi_k\right)\right]} + (1-\varphi_{\ell t})S_{\ell jt-1}.$$

Individuals who reoptimize have relatively standard logit choice probabilities (although we allow for income based hetereogeneity in the variance of the composit error). Individuals who do not reoptimize stay in the same practice as they were in $t-1$.

[18]We let $\tilde{\sigma}^2_{\nu_\ell} = \frac{6}{\pi^2}\sigma^2_{\nu_\ell}$.

5.1 Estimation Approach

We estimate the parameters governing preferences, consumer information, and inertia, $\Phi = \{\beta, \gamma, \theta, \xi\}$ using an indirect inference approach (Gourieroux et al. 1993). Our implementation, which incorporates moments from our regression discontinuity approach, is related to recent work by Fu and Gregory (2019) and Allende (2022) incorporating RD estimates into an empirical model of post-disaster subsidies and school choice, respectively. Our approach operates in two steps.

Step 1: Auxiliary Model

In the first step, we construct an auxiliary model using the specification described above. Given a guess of parameters Φ (and lagged shares $S_{\ell jt-1}$), we recover the model implied market shares $\tilde{S}_{\ell jt}(\Phi)$. We then use these market shares, and the size of each market, to construct enrollment and quarter to quarter enrollment changes. With enrollment changes for each GP$\times$LSOA pair, we are then able to estimate a regression discontinuity following the approach laid out in section 4.1. For computational simplicity we use a uniform kernel, as shown in Equation 5. This provides estimates of the jump at the threshold, as we well as the slopes on either side of the threshold. We run this separately for high and low income LSOA, and summarize the coefficients as a vector $\tilde{\tau}_I(\Phi)$, for $I \in \{\text{high, low}\}$.[19] We use the same sample restrictions used for the results in Section 4.3.

Step 2: Moment Conditions

In the second step, we use the output of our auxiliary model to construct two sets of moment conditions. The first set matches model implied market shares to realized market shares:

$$M^1(\Phi) = \mathbb{E}[\log(\tilde{S}_{\ell jt}(\Phi)) - \log(S_{\ell jt})] = 0 \quad \forall\, l, j, t. \tag{9}$$

Since there are no prices that could generate endogeneity concerns, estimation does not require instruments. The second set of moments matches our model implied regression discontinuity estimates to corresponding estimates from the data. Letting $\hat{\tau}_I$ represent the vector of slope and intercept coefficients estimated from the data (with $\hat{\tau}_{k,I}$ representing the kth element of this vector), the moments are

[19] In practice, estimates from the parametric and non-parametric RD model are similar. See Appendix Table A-5.

$$M^2(\Phi) = \mathbb{E}[\tilde{\tau}_{k,I}(\Phi) - \hat{\tau}_{k,I}] = 0 \quad \forall\, k,\, I \in \{\text{high,low}\}. \tag{10}$$

We then stack these moment conditions to form

$$M = \begin{bmatrix} M^1(\Phi) \\ M^2(\Phi) \end{bmatrix} = \mathbf{0}.$$

Our estimator is the solution:

$$\widehat{\Phi} = \text{argmin}_{\Phi} M' W M.$$

where W is a positive definite weighting matrix. We compute standard errors by numerically computing $\frac{\partial M}{\partial \Phi_k}$ for all elements of Φ.

Identification of Preferences and Information Precision

The use of the auxiliary model is important for identification. In general, it is difficult to separately identify unobserved quality (ξ_j), the preference for quality as measured by the reviews, and the precision of individuals' priors. Without targeting the RD estimates, ξ_j and the parameters related to reviews would only be identified based on GPs that cross the review ratings threshold, similar to the identification in the panel regression approach in Section 4.4. The RD moment exploits the fact that GPs on either side of the threshold have similar underlying quality. Figure 2 demonstrates that the slope of demand around the threshold for star ratings and the response at the threshold help identify the precision of individual's prior and the preference for quality. In particular, if either the slope or response at the threshold is positive, then individuals have a preference for quality. The response at the threshold relative to the slope provides information on the precision of an individual's prior. Furthermore, the difference between the RD estimates for low- and high-income patients helps identify γ_1, which governs heterogeneity in the precision of private information.

5.2 Model Estimates

For estimation, we focus on the sample of individuals in Greater London given computational constraints. The RD results for this sample are similar to the RD estimates for all of England (see columns 5 and 6 of Appendix Table A-5). As in the baseline results, low income individuals respond more at the star rating thresholds than high income individuals. In addition, demand for high income individuals is more responsive to the underlying average reviews away from the threshold, consistent with high income individuals already being informed about quality. We define the choice set, $\in \mathcal{J}_{lt}$, as the set of GPs within 10km of each LSOA l in quarter t.

We present the estimates from our structural model in Table 3. Our key focus is on two sets of parameters: those governing the variance of the private signal and preferences for GP quality. As in our reduced form evidence, our structural estimates suggest that high income patients have significantly more precise information about GP practice quality. Furthermore, both high and low income patients have strong preference for quality, as captured by r_{jt}, although this preference is marginally higher for high-income patients.

Table 3
Estimates for GP Demand Model

	Estimate	SE
Inertia		
Constant	−4.7864	(0.2295)
Income	0.0373	(0.0058)
Age	−0.0179	(0.0012)
Health	0.0977	(0.0472)
Private Signal Variance ($\sigma^2_{\ell t}$)		
Constant	−1.2526	(0.0694)
Income	−0.9932	(0.0772)
GP Quality		
Constant	0.9002	(0.0071)
Income	0.0382	(0.0044)
Distance		
Constant	−1.3286	(0.1779)
Income	−0.0503	(0.0043)
Age	0.0080	(0.0010)
Health	−0.0376	(0.0363)
Other GP Characteristics		
Mean physician age	−0.0384	(0.0015)
Practioners per patient	0.0008	(0.0000)
Active choice fraction	0.0066	
N	670,861	

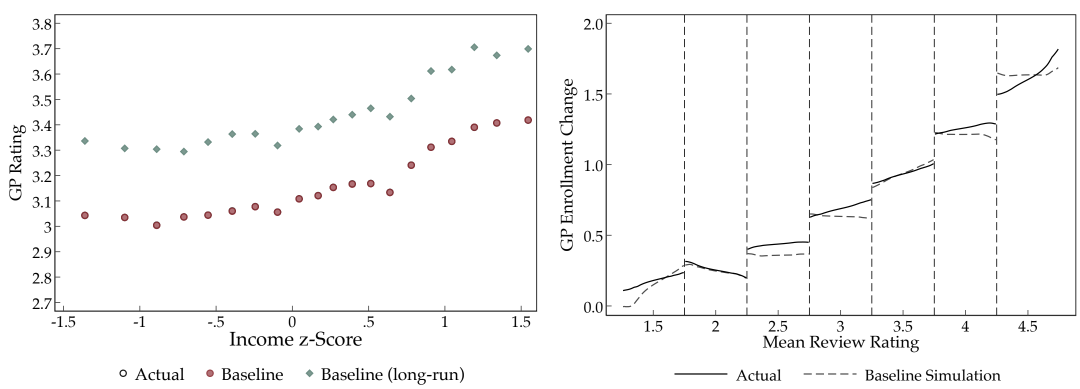
Notes: Estimates from demand model estimated via method of moments.

The parameters governing inertia and preferences for other observables are reasonable. For example, patients dislike GP practices that are far from the LSOA, and this preference is slightly stronger for high income LSOAs. The magnitude of the parameter governing switching is large, indicating that a small fraction of individuals switch each quarter. We estimate that 0.7% of individuals switch each quarter. We use these parameters to construct counterfactuals in the next subsection. These allow us to consider the short and long run impacts of various policy proposals.

5.3 Counterfactuals and Model Fit

We begin by considering the fit of our model. The hollow dots in Panel a of Figure 6 plot the relationship between LSOA income and GP practice quality in the greater London subsample we use to estimate our model. We overlay red dots, which represent the same relationship as recovered from our model. The two are virtually identical, suggesting that our model effectively captures market shares in our data. This is perhaps unsurprising, given the inclusion of GP practice fixed effects and lagged market shares, and the relatively small number of re-optimizing patients each period.

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 (ii) our model, and (iii) the long run implied by our model (allowing all individuals to switch). Panel b plots enrollment changes within rounded star rating bins both based on the data and implied by our model.

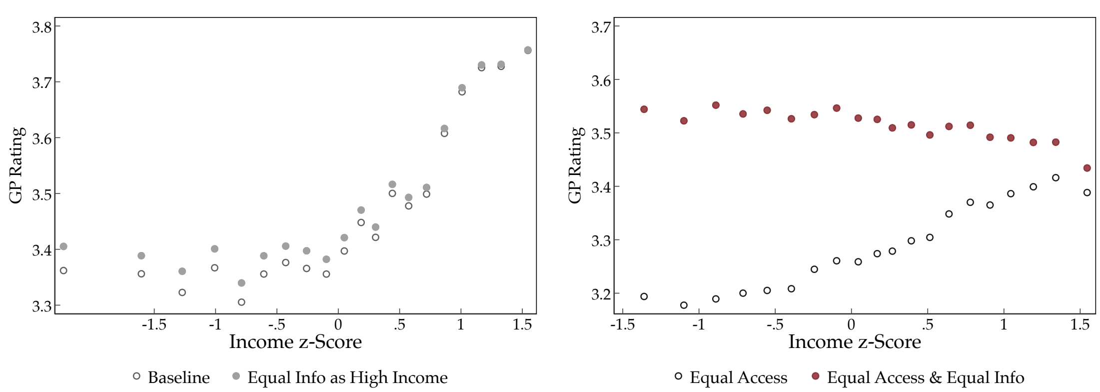
Panel b of Figure 6 shows that our model is also able to effectively capture the patterns with respect to *changes* in enrollment that we feature in our regression discontinuity. This chart repeats the analysis in Figure A-5, showing local linear smoothings of enrollment changes within each rounded star rating for (i) the greater London subsample (solid line) and (ii) the implied enrollment changes from our baseline model simulation. As in the data, our baseline simulations show sharp responses at each star rating threshold.

Figure 6 also shows a first simple counterfactual enabled by our model. The gray diamonds in panel a show the long run quality disparities implied by our model. We simulate these outcomes by setting the re-optimization probability $\varphi = 1$ and fixing all other parameters as in our baseline simulations. We see a notable level shift across the income distribution: as both high and low income patients re-optimize, they tend to choose higher quality practices. However, the gradient between income and GP quality is remains roughly constant despite this level shift. We now turn to considering our two main counterfactuals.

Equating Information

Our first counterfactual considers the impact of equalizing information on the relationship between income and GP practice quality. To do so, we re-simulate our model, but set the precision of all patients' private information to the precision of the highest income ventile. This allows us to consider fraction of the observed income-quality gradient that can be explained by information itself.

Figure 7
Counterfactual Experiments



a. Equating Information

b. Equating Information and Access

Notes: Panel a shows baseline relationship between income and quality and counterfactual relationship setting precision of information for all patients equal to precision for high income LSOAs. Panel b shows relationship equating access (randomly assigning quality for all GP practices) and additionally equating information as in panel a.

Panel a of Figure 7 shows the results of this experiment, which suggests that information is responsible for a non-trivial portion of disparities by income. The hollow-dots show the long-run choices under our baseline parameters. The gray dots show the long-run choices after equating the precision of information for all patients. There is a noticeable jump in average quality chosen by lower income patients, declining across the distribution, with no change for the top ventile (who have the same information in both scenarios). We summarize the results of this counterfactual in Table 4. We find that equating information reduces the gradient between income and quality by roughly 8 percent. While substantial disparities remain, this is to be expected given meaningful differences in access (and potentially in preferences as well).

Table 4
Summary of Counterfactuals

	Baseline		Equal Info		Percent Change
	Low Income	High Income	Low Income	High Income	in Correlation
Mean distance to GP (km)	0.69	0.90	0.69	0.90	−0.80
Mean rating	3.33	3.71	3.36	3.72	−7.74
Mean QOF clinical	−0.12	0.14	−0.11	0.14	−1.70

Notes: Summary of Counterfactual Exercises. Baseline columns refer to long-run simulations under our baseline parameter estimates. Equal info columns refer to long-run simulations setting the precision of information equal across all participants. Percent change in correlation refers to the change in the unconditional correlation between LSOA income and r_{jt} across the two simulations.

Equating Access and Information

The fact that a substantial relationship between income and quality remains even after equating information does not necessarily mean that access is a key driver. It is possible, for example, that the remaining disparities are driven by differences in preferences by income. To assess this possibility, our final set of counterfactuals considers the impact of equating both access and information.

There are multiple potential ways to equate access. Our approach fixes the set of available GPs in the choice set for each LSOA but randomizes the quality r_{jt} for each practice by drawing with replacement from observed quality in the full distribution. We then run two versions of our simulation. In the first, we allow information to vary in the population as in our baseline model. In the second, we fix the precision of all participants to the precision of the top ventile, as in our earlier counterfactual.

The results, shown in panel b of Figure 7 show that equating access is insufficient to eliminate the income-quality gradient, but that equating both access and information is sufficient. The hollow dots present the simulated gradient equating only access. Given the importance of information, we still see an upward slope. The red dots present the simulated gradient equating both access and information. This elimates the gradient entirely, and even generates a slight negative relationship, likely because high income patients are marginally more willing to trade-off quality for distance. In other words, the source of observed disparities does not appear to be preferences.

Of course, a key caveat to these counterfactuals is that they assume a fixed supply-side in the long run. It is entirely possible that capacity constraints or the endogenous entry, exit, or other responses of gp practices could alter our conclusions. However, the consequences of a supply-side response for the gradient itself are ex-ante unclear. While a fully developed general equilibrium model is beyond the scope of this paper, our counterfactuals provide a benchmark from which to consider richer models that incorporate the incentives of GP practices.

6 Conclusion

While a large literature has documented disparities in the quality of health care received by patients, the role of information has been largely ignored. Focusing on GP practice choice in England, we find evidence consistent with high-income individuals having better information about provider quality. Regression discontinuity evidence suggests that residents of low-income neighborhoods are more responsive to the information provided by star ratings shown on the NHS website, despite both low and high income patients valuing quality.

We estimate a structural model based in which individuals have heterogeneous private information about provider quality. The model can rationalize observed differences in patient responses to star ratings. We use the model to simulate how choices would change if information was equated across the population. Counterfactuals imply a significant reduction in the income-quality gradient, implying that differential information plays an important role in driving inequality in health care quality. However, it is important to note that differential access is still an important factor, and one that might be further exacerbated by supply-side responses to changes in the informational environment.

References

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

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

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.

Bajari, Patrick, Han Hong, Minjung Park, and Robert Town, "Regression discontinuity designs with an endogenous forcing variable and an application to contracting in health care," Technical Report, National Bureau of Economic Research 2011.

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

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.

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.

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.

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.

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.

Dynarski, Susan, CJ Libassi, Katherine Michelmore, and Stephanie Owen, "Closing the gap: The effect of reducing complexity and uncertainty in college pricing on the choices of low-income students," *American Economic Review*, 2021, *111* (6), 1721–56.

Dynarski, Susan M and Judith E Scott-Clayton, "The cost of complexity in federal student aid: Lessons from optimal tax theory and behavioral economics," *National Tax Journal*, 2006, *59* (2), 319–356.

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

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

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

Handel, Benjamin R, "Adverse selection and inertia in health insurance markets: When nudging hurts," *American Economic Review*, 2013, *103* (7), 2643–82.

__ , 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.

Ho, Kate, Joseph Hogan, and Fiona Scott Morton, "The impact of consumer inattention on insurer pricing in the Medicare Part D program," *The RAND Journal of Economics*, 2017, *48* (4), 877–905.

Hortaçsu, Ali, Seyed Ali Madanizadeh, and Steven L Puller, "Power to choose? An analysis of consumer inertia in the residential electricity market," *American Economic Journal: Economic Policy*, 2017, *9* (4), 192–226.

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.

Kling, Jeffrey R., Sendhil Mullainathan, Eldar Shafir, Lee C. Vermeulen, and Marian V. Wrobel, " Comparison Friction: Experimental Evidence from Medicare Drug Plans *," *The Quarterly Journal of Economics*, 01 2012, *127* (1), 199–235.

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.

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.

Luco, Fernando, "Switching costs and competition in retirement investment," *American Economic Journal: Microeconomics*, 2019, *11* (2), 26–54.

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.

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

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

Oreopoulos, Philip and Reuben Ford, "Keeping college options open: A field experiment to help all high school seniors through the college application process," *Journal of Policy Analysis and Management*, 2019, *38* (2), 426–454.

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.

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.

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.

Urquiola, Miguel and Eric Verhoogen, "Class-size caps, sorting, and the regression-discontinuity design," *American Economic Review*, 2009, *99* (1), 179–215.

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.

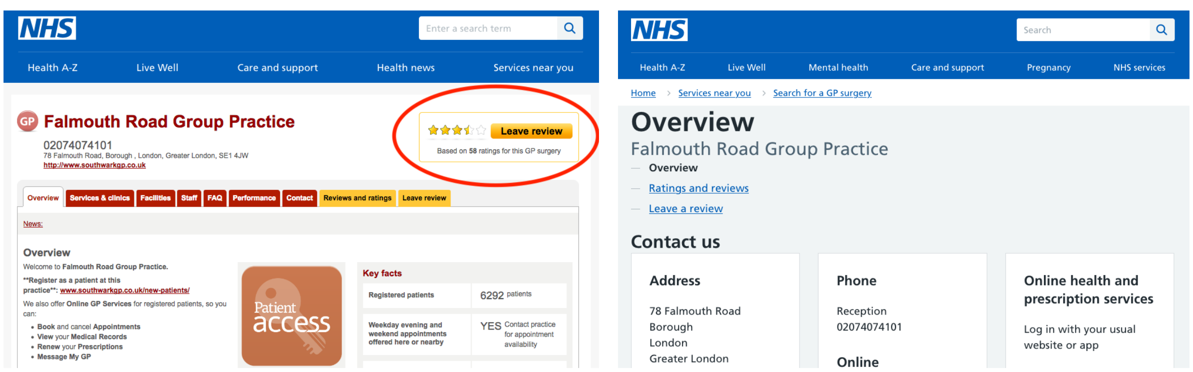
__ , Xander Koolman, and Andrew M Jones, "Explaining income-related inequalities in doctor utilisation in Europe," *Health economics*, 2004, *13* (7), 629–647.

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.

Online Appendix

A Appendix Figures

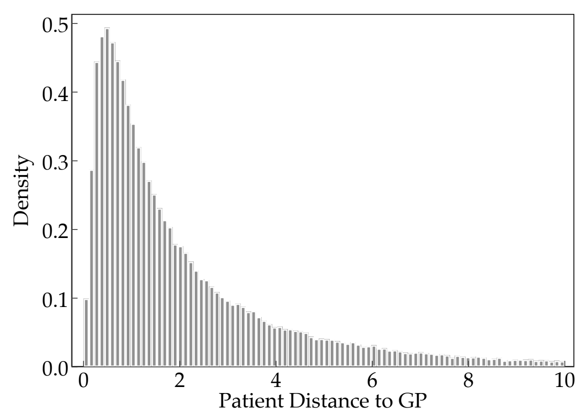
Figure A-1
Example of the NHS Website Before and After Website Change



a. Visible star ratings b. No star ratings

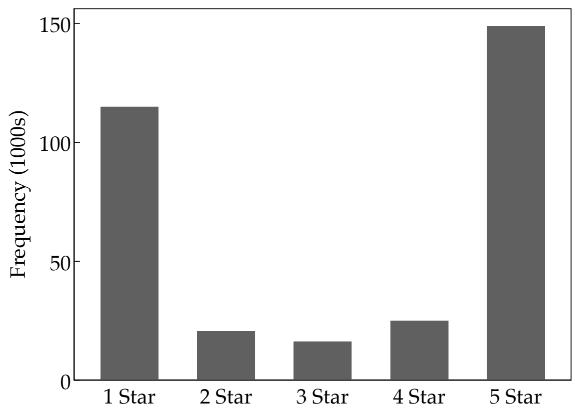
Notes: Shows NHS website for a GP practice prior to January 2020 (with visible star ratings) and after January 2020 (with no visible star ratings).

Figure A-2
Histogram of Distance to GP



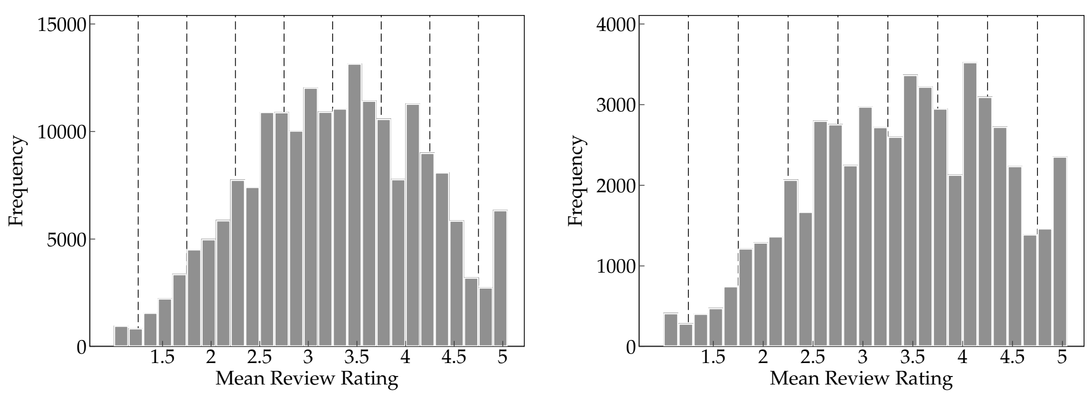
Notes: Chart shows histogram of individual distance to GPs.

Figure A-3
Histogram of Individual Reviews



Notes: Chart shows histogram of individual reviews for May 2012 to December 2021.

Figure A-4
Histogram of Average GP Reviews

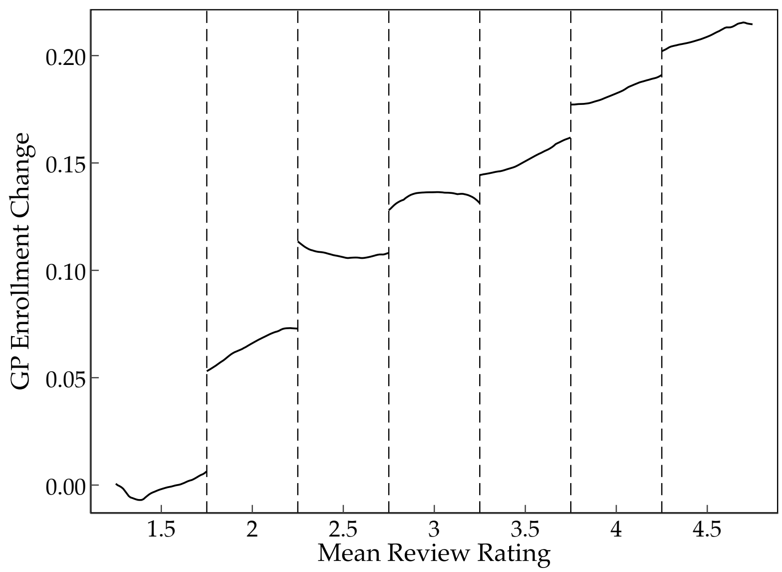


a. Visible star ratings

b. No star ratings

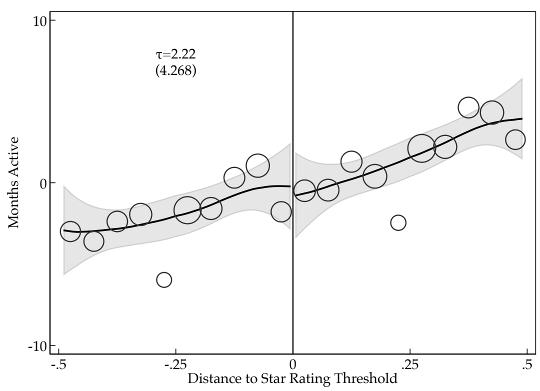
Notes: Chart shows histogram of average reviews for the period when the NHS Choices website displayed star ratings ("visible star ratings") and the period when the website did not display star ratings ("no star ratings"). The NHS calculated average reviews using the running average of individual reviews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

Figure A-5
GP Enrollment Change and Review Thresholds

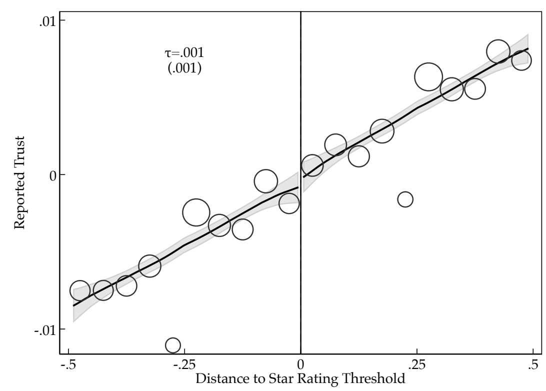


Notes: Chart shows the relationship between average reviews and GP enrollment change by GP-LSOA-quarter for the period when the NHS Choices website displayed star ratings. Lines are smoothed using a local linear regression. Vertical lines show thresholds for rounded star ratings.

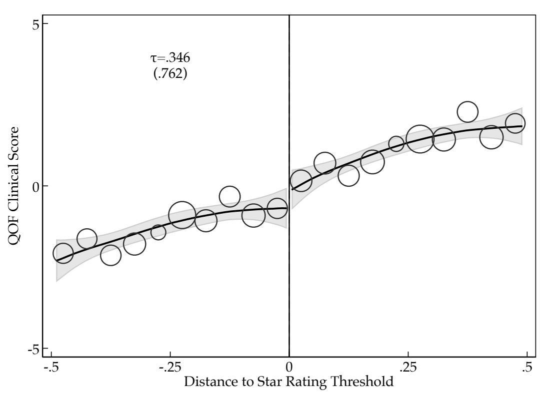
Figure A-6
Smoothness of GP-level Covariates Around Threshold



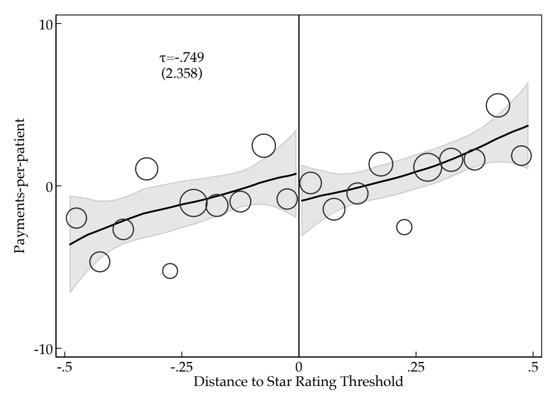
a. Months Active



b. Confidence and Trust (Survey)



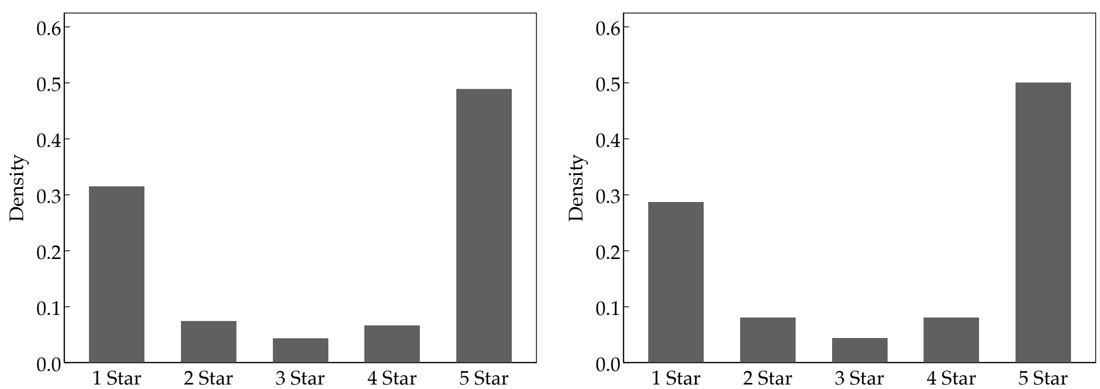
c. QOF Clinical Score



d. Payments Per-Patient

Notes: Changes at the threshold in GP-level observables. Circle size corresponds to the number of observations in each bin. Fitted line is from a local linear regression with a triangular kernel. Shaded area shows 95% confidence intervals.

Figure A-7
Distribution of Individual Reviews Around Website Change

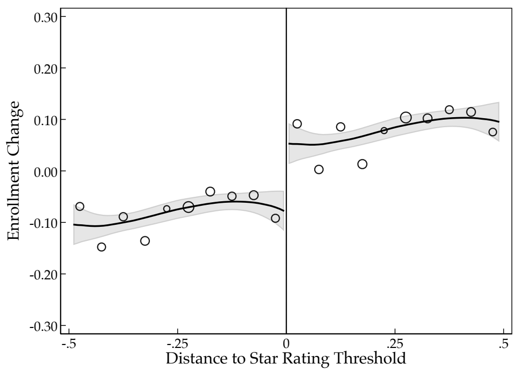


a. Visible Star Ratings

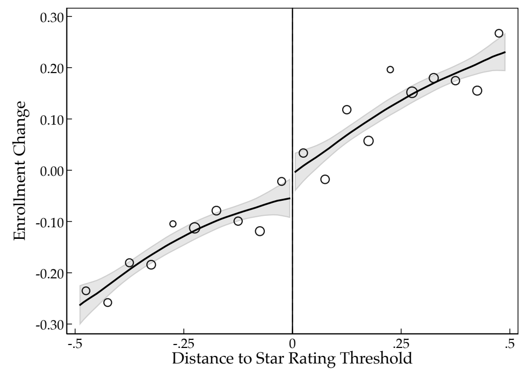
b. No Star Ratings

Notes: Sample is individual reviews a month before and after the website change in January 2020. The Chi-squared value is 5.6556 (p-value 0.226).

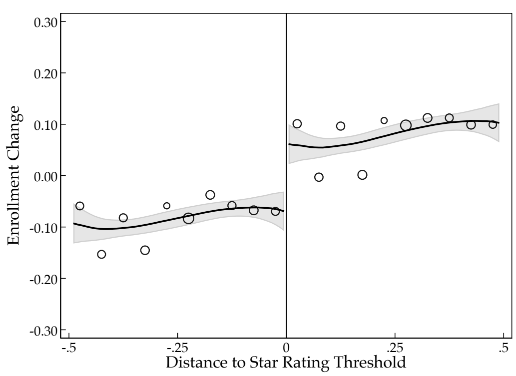
Figure A-8
Effect of Star Rating Threshold on Enrollment Change
Additional Heterogeneity Analysis



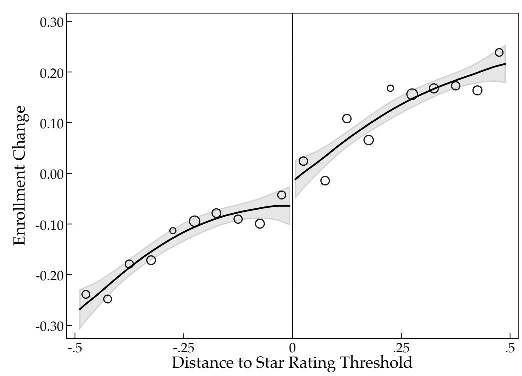
a. Low Education



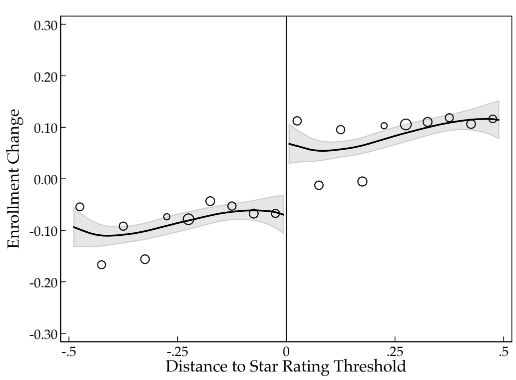
b. High Education



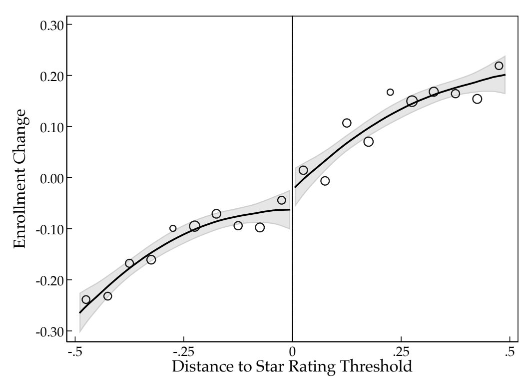
c. Low Employment



d. High Employment



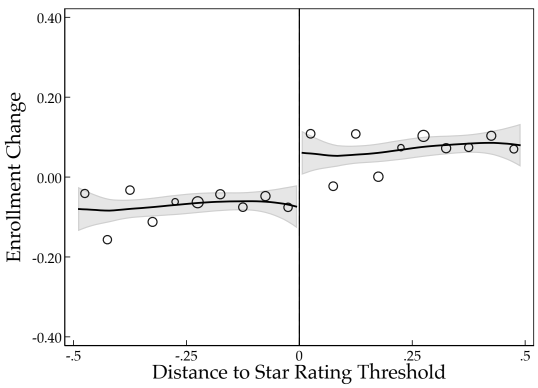
e. Low Health



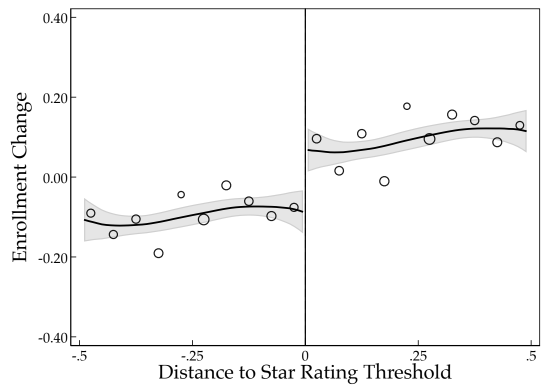
f. High Health

Notes: Chart shows mean enrollment change around threshold for star ratings by quartile of LSOA income of patients for the period in which star ratings were visible. The size of the circles corresponding to the number of observations in each bin. The fitted line is from a local linear regression using a triangular kernal. Shaded area shows 95% confidence interval.

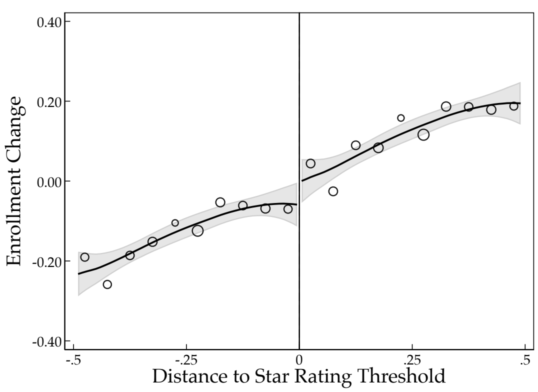
Figure A-9
Effect of Star Rating Threshold on GP Enrollment
by Income Quartile



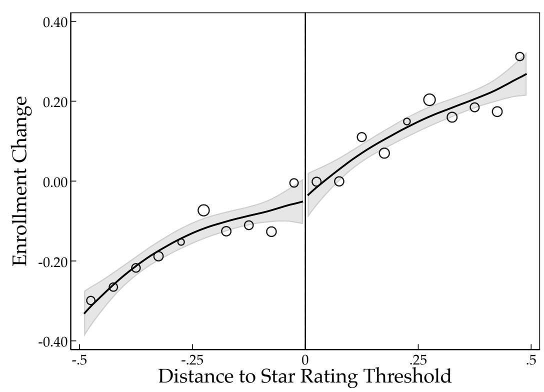
a. First Quartile
(Lowest Income)



b. Second Quartile



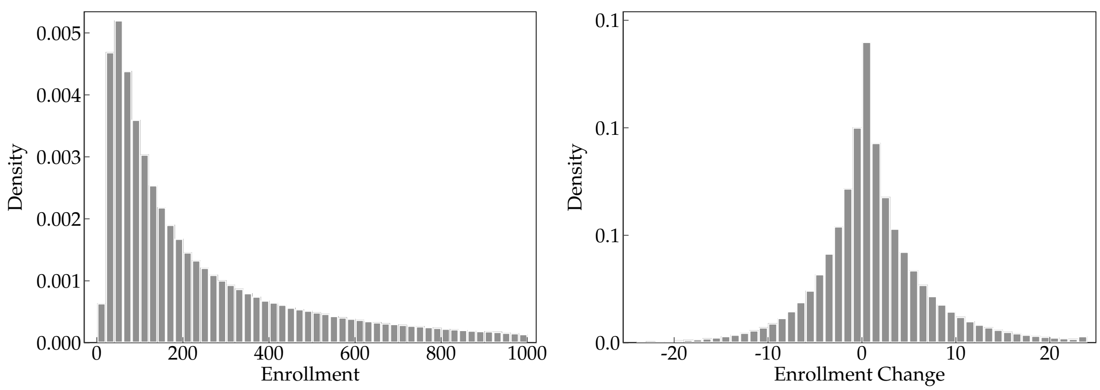
c. Third Quartile



d. Fourth Quartile
(Highest Income)

Notes: Chart shows mean enrollment change around threshold for star ratings by quartile of LSOA income of patients for the period in which star ratings were visible. The size of the circles corresponding to the number of observations in each bin. The fitted line is from a local linear regression using a triangular kernal. Shaded area shows 95% confidence interval.

Figure A-10
Histogram of GP Enrollment by LSOA-Quarter

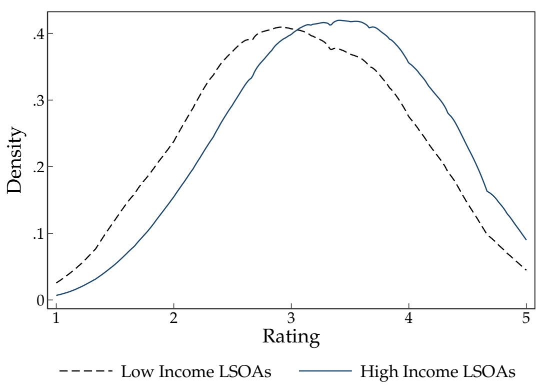


a. Level

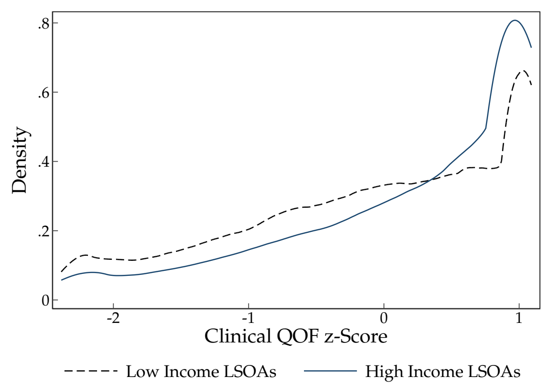
b. Change

Notes: Chart shows histogram of enrollment by LSOA-quarter and quarterly change in enrollment. Enrollment above 1,000 not shown in panel (a). Absolute change in enrollment greater than 25 not shown in panel (b).

Figure A-11
Density of GP Quality in Choice Set by Income



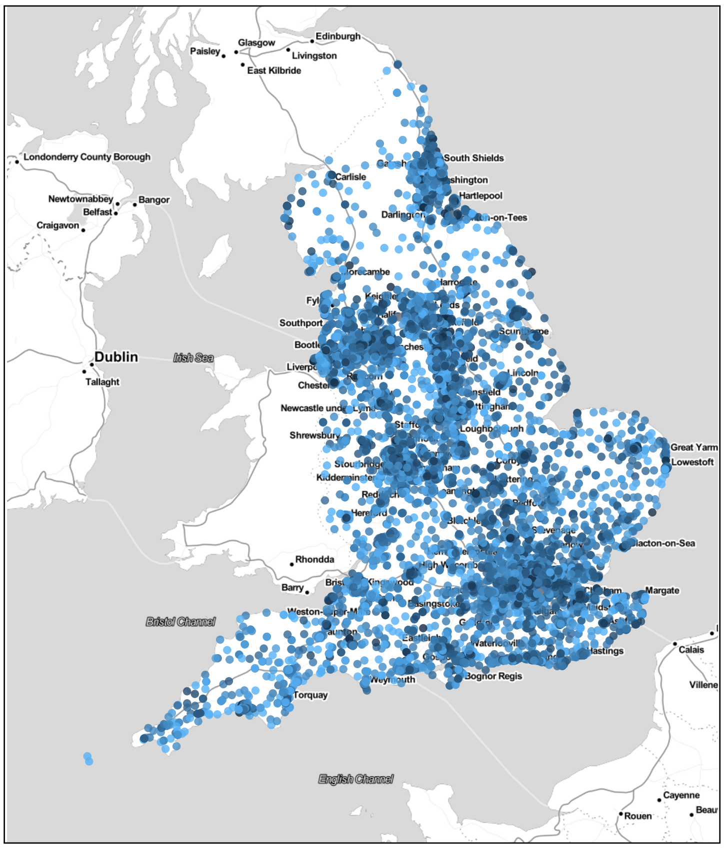
a. Subjective Ratings



b. Clinical Quality Outcomes Framework

Notes: Charts shows histogram of GP ratings and GP clinical Quality Outcome Framework scores for GPs in individuals' choice set, defined as all GPs within 5km. Low (high) income LSOAs are defined as those with an income in the first (fourth) quartile. GPs with less than 5 reviews are excluded from the left chart.

Figure A-12
Location of GPs and Enrollment



Notes: Map shows the location of GPs. Darker color corresponds to higher enrollment.

B Appendix Tables

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: Shows correlation coefficient between relevant variable and mean patient review along with the relevant p-value.

Table A-2
Robustness for Full Sample RD Estimates

	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, 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, controls for GP age, age squared, and number of practitioners in the GP practice, as well as threshold fixed effects are included, and MSE-optimal bandwidths are used. High and low income refer to above vs. below median income deprivation at the LSOA level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A-3
Robustness for RD Estimates by Income

	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: Alternative Bandwidths									
	Bandwidth=0.1		Bandwidth=0.2		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	0.223***	0.095	0.186***	0.034	0.145***	0.027	0.119***	0.027	0.111***	0.024
	(0.084)	(0.082)	(0.057)	(0.055)	(0.047)	(0.045)	(0.042)	(0.040)	(0.038)	(0.037)
Bandwidth	0.10	0.10	0.20	0.20	0.30	0.30	0.40	0.40	0.50	0.50
N	347,252	351,372	708,700	722,588	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 bandwidth choice. 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, controls for GP age, age squared, and number of practitioners in the GP practice, as well as threshold fixed effects are included, and MSE-optimal bandwidths are used. High and low income refer to above vs. below median income deprivation at the LSOA level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A-4
Effect of Star Ratings on Enrollment Change
Panel Regression Estimates

	(1)	(2)	(3)	(4)
Stars $\times$ 2	0.029∗∗∗	0.025∗∗∗		
	(0.001)	(0.001)		
(Stars$\times$2) $\times$ 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)$\times$ 1(Low Income)				0.113∗∗∗
				(0.012)
1(Stars=2)$\times$ 1(Low Income)				0.115∗∗∗
				(0.007)
1(Stars=2.5)$\times$ 1(Low Income)				0.100∗∗∗
				(0.006)
1(Stars=3)$\times$ 1(Low Income)				0.073∗∗∗
				(0.008)
1(Stars=3.5)$\times$ 1(Low Income)				0.045∗∗∗
				(0.005)
1(Stars=4)$\times$ 1(Low Income)				0.029∗∗∗
				(0.006)
1(Stars=4.5)$\times$ 1(Low Income)				0.006
				(0.008)
1(Stars=5)$\times$ 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.

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

			Visible Star Ratings		Visible Star Ratings London Only	
	Visible Star Ratings	No Star Ratings	Low Income	High Income	Low Income	High Income
Estimate	0.072	0.017	0.113	0.035	0.159	0.047
	(0.028)	(0.049)	(0.035)	(0.034)	(0.064)	(0.063)
Distance from threshold	−0.031	0.099	−0.129	0.065	−0.216	0.039
	(0.056)	(0.097)	(0.069)	(0.068)	(0.129)	(0.127)
Outcome Mean	0.82	0.41	0.87	0.77	0.93	0.90
N	3,421,544	1,116,437	1,698,686	1,722,858	564,239	498,160

Notes: Dependent variable is change in GP enrollment. Sample excludes quarterly GP observations with fewer than five reviews. Each regression controls linearly for average review score above and below the threshold. Includes star ratings 1.5 to 4.5. Standard errors are in parentheses and clustered by GP.