

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

July 2022

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 provider quality is an important determinant of this disparity. Our empirical strategy exploits the temporary presence of a website that publicly displayed summary star ratings of general practitioner (GP) offices in England. Regression discontinuity (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 our RD moments into a structural model of demand that allows for heterogeneity in information in addition to consumer inertia and heterogeneous preferences. Our results indicate that a meaningful fraction of the income-quality gradient could be eliminated by removing informational differences.

*We are grateful to Martin Gaynor, Ben Handel, Maria Polyakova, Carol Propper, and seminar participants at Berkeley, Caltech, the CEPR virtual IO seminar, Michigan, NBER Summer Institute, Tsinghua, and the ZEW Conference on Information and Communication Technologies for valuable comments. We thank Radoslaw Kowalski for sharing data.

1 Introduction

There are persistent within-country inequalities in health care throughout the world (Cookson et al. 2021; Hart 1971). Even in countries with free public health services, like the United Kingdom, lower income individuals receive lower quality care (van Doorslaer et al. 2004, 2006; OECD 2014; Cookson et al. 2016; Scobie and Morris 2020).¹ A key question for policy-makers is whether these disparities are solely driven by access to services, or whether they reflect differences in information or preferences about health care providers.

In this paper we show that differential information is an important cause of health care inequality, i.e., of the positive relationship between patient income and general practitioner (GP) quality in the UK. Our empirical strategy focuses on the temporary presence of a website run by the UK National Health Service (NHS) intended to provide easily accessible information about GP quality. The website, known as *NHS Choices*, publicly displayed GP quality ratings of one to five stars rounded to the nearest half-star. We implement a regression discontinuity (RD) 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 RD estimates show that patients respond sharply to the information provided by the website. While the website was active, GP practices with higher star ratings experienced greater enrollment growth, compared to other GPs with nearly identical values of the underlying quality index. Crucially, this increase in demand is almost entirely driven by individuals in low-income neighborhoods. While individuals in high-income neighborhoods are sensitive to provider quality in general, they do not appear to respond to the information contained in the website's star ratings. This is consistent with an information gap: well informed (high-income) individuals whose priors are not shifted by the website's information, and less-well informed (low-income) individuals, whose beliefs about quality are strongly influenced by the star ratings.

To assess the importance of these information gaps in explaining broader health-care inequalities, we incorporate our RD estimates into a structural model of GP choice. We allow for inertia and heterogeneity by income in both preferences and the precision of patient

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

information about GP quality. Counterfactual simulations imply that a substantial fraction of the relationship between income and healthcare quality is due to this information gap. These results suggest that relatively low-cost informational interventions can significantly 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, comprehensive data on GP registration is available, allowing us to examine enrollment decisions for all GP practices in England using detailed geographic information about patients. This is combined with a broad range of quality metrics for GPs, from both survey-based patient evaluations and objective measurements of clinical outcomes. Second, the government-sponsored universal health care 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, in January 2020 the NHS website stopped prominently displaying summaries of the star ratings (but continued to collect reviews). 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.² Consistent with previous work, individuals have a preference for quality GPs (Santos et al. 2017). We show that, across different measures of quality, individuals who live in higher-income areas tend to choose better GPs. A portion of this can be explained by access: GPs located in high-income areas tend to be of higher quality. However, this cannot explain the entirety of the relationship. Even conditional on the choice set, individuals in high-income areas tend to select higher quality GPs. This pattern is consistent with high-income individuals having more detailed information about provider quality. However, there are alternative explanations that could rationalize observed patient choices, including unobserved barriers (e.g. discrimination), reverse causality (e.g., performing well on various quality metrics is easier when serving a higher income population), and preferences of patients over attributes not captured by observed quality measures.

To guide our analysis, we next present a theoretical framework in which individuals

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

combine private information with a publicly observed rounded star rating to form beliefs about GP quality. The model delivers sharp predictions about the relationship between ratings and patient choice that motivate our regression discontinuity approach. Choices of well informed patients will be strongly influenced by true quality, but relatively unresponsive to coarse star ratings. Choices of less informed patients will respond sharply to star ratings, but are only weakly related to true quality otherwise. This framework also forms the basis of the structural model of demand used to quantify the role of information disparities.

Our RD 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 2007 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 be given a star rating of 4 stars, while a provider with an index value of 3.74 will appear with 3.5 stars.³ Following an approach used in the previous literature, we compare demand just above or below these discontinuities (e.g., Anderson and Magruder 2012; Luca 2016).

Our results show that patients respond to the information contained in 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 at the end of 2019, this difference disappeared.

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, low-income patients are uncertain about provider quality and benefit 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

³We 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.

growth is highly correlated with the underlying index for high-income areas, and effectively uncorrelated for low-income areas. Again, 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.

We examine the robustness of the results in several ways. We find similar results when restricting our sample to movers—patients who relocate geographically and are therefore extremely like to switch their GP. This confirms that our findings are not driven by differential inertia by income level. We also find similar results when implementing a panel regression strategy which focuses on practices that experience changes in the summary star ratings over time.

As a final step, we construct and estimate a structural model of GP choice. 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 the extent to which information plays an economically meaningful role in observed health care disparities. Our structural 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. Finally, we account for the potential capacity constraints of GPs by making demand a function of patients per practitioner.

We estimate the model using an indirect inference approach (Gourieroux et al. 1993) that incorporates our RD results in addition to matching shares. Targeting the RD estimates in the estimation procedure helps disentangle heterogeneous information and heterogeneous preferences as drivers of GP demand. Estimates from the demand model imply that high income individuals have more precise information about quality whereas differences in preferences for low and high income individuals is not a primary driver of differences in demand.

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 equalizing the information available to high- and low-income patients can remove a substantial fraction of the relationship between income and GP quality, on the order of 8%. Furthermore, the entire relationship disappears once we equalize both information and access to quality GPs. This suggests that differences in preferences are not an important driver

of the observed income-quality gradient. In addition, the results imply that the interaction of access and information is important. Increasing access to quality alone will not benefit low-income patients who are not aware of the higher quality providers. 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 inequality in health provider choice. This serves as a bridge between two large literatures. The first is a broad body of work highlighting the existence and persistence of healthcare inequalities by income (e.g. Hart 1971; Peters et al. 2008; van Doorslaer et al. 2006; Gwatkin et al. 2004; Cookson et al. 2016; Marmot et al. 2007; Balarajan et al. 2011; Devaux 2015), 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 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). There is also evidence that physicians and their families make different health care decisions, including drug choice and preventative care, consistent with having more health care expertise (Bronnenberg et al. 2015; Chen et al. 2019; Artmann et al. 2019). Our innovation comes in highlighting information as an important source of disparities in health care provider choice.

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), Anderson and Magruder (2012), Lewis and Zervas (2016), Luca (2016), Luca and Zervas (2016), Reimers and Waldfogel (2021) and more, particularly those that explicitly incorporate learning models (e.g. Newberry and Zhou 2019). Recent work has examined the effect of review systems on doctor congestion and interruptions in care (Chartock 2021; Kummer et al. 2021). Our primary contribution is in using the response to reviews to identify information disparities.

The remainder of this paper is as follows. Section 2 provides additional background on the website and data. Section 3 presents a stylized model of GP practice choice in the presence of uncertainty and a rating website. Section 4 introduces and present results from our RD design. Section 5 presents our structural model 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, including details on the government website providing information about GPs. 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 doctors, which contract directly with the National Health Service (NHS).⁴

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.⁵ Absent capacity

⁴Most GP funding is based on the number of patients enrolled at the practice, weighted by patient complexity. This is supplemented with performance pay based on a number of schemes such as the Quality and Outcomes Framework (QOF), if a GP voluntarily participates.

⁵The constitution states “You have the right to choose your GP practice, and to be accepted by that prac-

constraints or a set of specialized circumstances, practices must generally accept all patients who wish to register.⁶ The process of registration is straightforward and can often be completed online. 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. The website provides information to help patients choose a GP practice. Moreover, some GPs allow patients to access their records, make appointments, and order repeat prescriptions through the website.

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.⁷ 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 in 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 removed the summary star ratings from the provider pages (see Figure A-2). While individuals could still theoretically find individual rankings and manually calculate the average, the information was significantly more difficult to access given that the review system was

6" unless there are reasonable grounds to refuse." See <https://www.gov.uk/government/publications/the-nhs-constitution-for-england>.

⁶<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.

⁷In 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.

de-emphasized. 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 star ratings were removed as a falsification exercise.

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 2016.⁸ The combined dataset contains all reviews from May 2013 to December 2021. We use these to construct a monthly panel from May 2015 to December 2021 that contains two key variables for each practice: the index r_{jt} , which is the precise average across all rankings (based on a two year moving window) of practice j in time period t 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 on 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 quality. 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. Also, r_{jt} is 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.⁹ This is consistent with previous evidence documenting a correlation with patient

⁸See Kowalski (2017).

⁹The 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

surveys and objective health outcomes, including evidence that provider ratings on NHS Choices are correlated with mortality and readmission rates (Greaves et al. 2012b,a).

The second concern is that there may be credibility issues, particularly if fake ratings 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.¹⁰ 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, including representative surveys, further helps assuage credibility concerns. In addition, because our primary empirical strategy is a RD design, the main threat to identification is the presence of differential fake ratings near rounding thresholds for summary star ratings. We consider this possibility and examine the density of reviews around the thresholds 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 analysis considers only GP practices with at least 5 reviews (6,673 practices satisfy this restriction at some point, or 87% of practices). 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

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

¹⁰See www.nhs.uk/our-policies/comments-policy.

each GP practice, we observe quarterly enrollment by age, gender, and Lower Layer Super Output Area (LSOA). LSOAs are detailed geographic areas with an average of about 700 households (2,000 individuals), and are roughly analogous to census block groups in the US. We 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 tend to remain enrolled 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. 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.¹¹

To supplement our analysis of enrollment change at the LSOA level, we obtained restricted individual-level enrollment records through a data access request to the NHS. These data provides information on the subset of patients that joined a GP practice in each month and whether they joined after moving to a new LSOA. The data contain an individual identifier, age, the month of a switch, the GP practice that an individual joined, the patient's most recent LSOA of residence, and whether the individual moved during the month. These data allow us to examine GP demand for the subset of individuals that were forced to actively choose a GP practice because they no longer live near their old GP. For this sample, we define income using the destination LSOA.

2.3 Summary Statistics

The first panel of Table 1 shows summary statistics for the GP rating system. The first two columns show data from the period in which summary star ratings were visible (until December 2019). Our unrestricted sample includes 7,635 unique GP practices, over 18 million

¹¹Although GP mergers are thought to be common, there is no explicit way to identify them in our data.

GP×LSOA×quarter observations and over 350,000 individual star reviews. During this period, the mean GP practice has an index of 3.2 stars. There is a large degree of dispersion of GP quality as measured by ratings. The standard deviation across practices is 1.0 stars and over 10 percent of GPs have 2 stars or less. The distribution of individual rankings and the average index value r_{jt} at the GP-month level are shown in Figure A-5 and Figure A-6, respectively. Most individual rankings are either 1 star or 5 star, however Figure A-6 shows that the distribution of the index across GP practices is relatively smooth, including near the star thresholds.

Summary statistics on GP practice enrollment and patient demographics are also shown in Table 1. 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%. Average enrollment was slightly higher in the post-star rating period, at just over 9,000. 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.¹² Appendix Figure A-4 shows a histogram of distance between LSOA centroids and chosen GP practices. Individuals tend to register with a GP within a short distance of their residence.¹³ The median individual lives only 1.4 km from their GP, suggesting distance is an important determinant of choice.

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 areas that are one standard deviation above the mean in terms of this income measure attend GP practices that are rated roughly 0.3 stars higher relative to those living

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

¹³See Figure A-15 showing the location of GPs in England.

Table 1
Summary of GP Enrollment and Characteristics

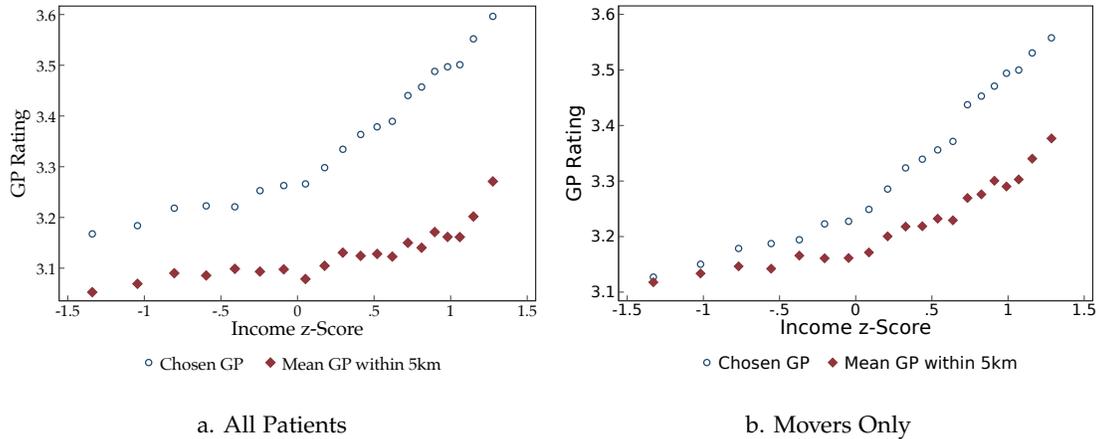
	Period with Star Ratings		Period without Star Ratings	
	Mean	SD	Mean	SD
<i>GP Reviews:</i>				
Individual review	3.17	1.84	3.43	1.69
GP average stars	3.20	1.02	.	.
GP Number of Reviews	84.5	89.7	122.6	146.0
<i>GP Enrollment:</i>				
Total Enrollment (100s)	80.75	50.92	91.38	60.50
LSOA Enrollment	0.58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

Notes: Table presents summary statistics for data on reviews and enrollment covering GPs in England from May 2015 to December 2021. Total enrollment data is at the GP-quarter level. LSOA enrollment is at the GP-LSOA-Quarter level. An LSOA is a small geographic area with about 700 households (2,000 individuals).

in LSOAs that are one standard deviation below the mean. This pattern is not an artifact of the star rating system. Appendix Figure A-3 shows the relationship between income and the Quality and Outcomes Framework clinical score, an objective measure of GP clinical 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 previous work documenting inequality in health care and preference for quality providers in the UK (e.g. Santos et al. 2017; O’Dowd 2020; Scobie and Morris 2020).

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 5 kilometers of the LSOA

Figure 1
Relationship between Income and GP Rating



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

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.

Figure 1 Panel b shows that the patterns in the data are even more stark when focusing on movers—individuals who move residences across LSOAs. There is a positive relationship between income (of the origin LSOA) and the quality of patients' chosen GP, and the relationship is stronger than the corresponding relationship between income and the choice set. Notably, movers at the low end of the income distribution choose GP practices that are roughly equivalent to the average within their choice set. This suggests that these individuals have very little information about practice quality in their new neighborhoods.

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 (e.g., 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 (differential) 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 private 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 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 $\epsilon_{ij} \sim N(0, \sigma_i^2)$. The precision of the signal, σ_i^2 , may differ across individuals given differential access to information or ability to conduct research. Given the prior and the signal, individuals form posterior beliefs using Bayes' rule. Expected quality is therefore given by

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

The weight on the private information, α_i , is given by

$$\alpha_i = \frac{\sigma_s^2}{\sigma_i^2 + \sigma_s^2}. \quad (2)$$

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 v_{ij} . Given these assumptions, expected utility can be expressed as:

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

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

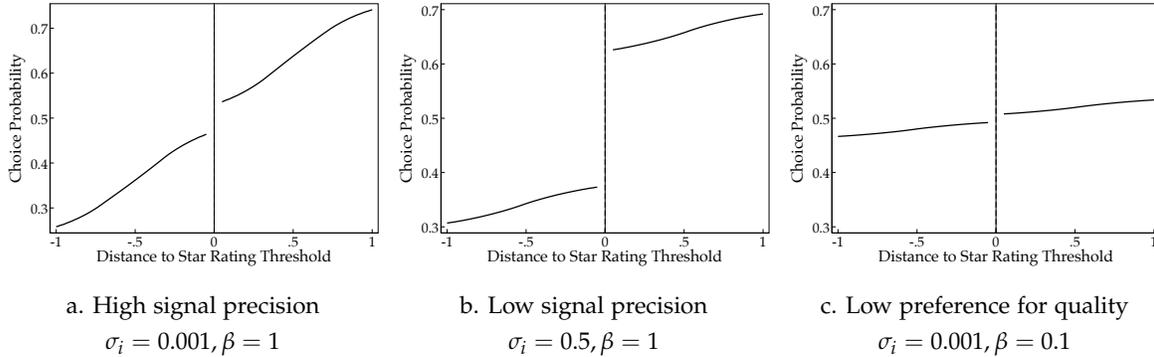
3.3 Implications for Ratings and Patient Choice

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 in which each individual has two providers in her choice set and observes two possible rounded star ratings: $s_j \in \{0, 1\}$.¹⁴ 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, individuals that can easily conduct research on the internet or 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 rating itself. This is reflected in the relatively modest jump at the star rating threshold.

¹⁴For 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.

Figure 2
Simulated Response to Star Rating Threshold

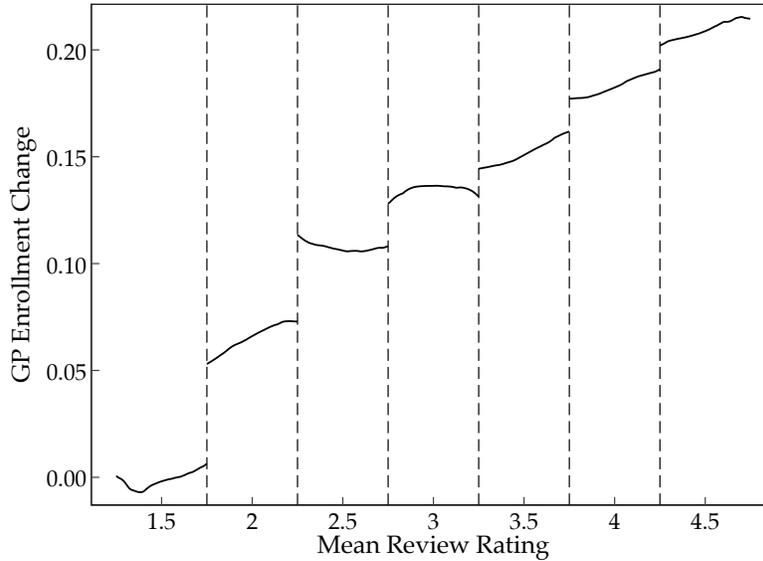


Notes: Charts show simulated demand as a function of quality when individuals receive a signal from star ratings. We assume individuals choose between two GPs with $r_j \sim N(0, 1)$. Stars are $s_j \in \{0, 1\}$, e.g. can be high or low depending on whether $r_j > 0$. In panel a, individuals have precise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

In contrast, individuals with less precise private signals, such as individuals 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.

Figure 3
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.

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).

Figure 3 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

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

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 variable, we consider $r_{jt}^0 = r_{jt} - c_s$, 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 $r_{jt}^0 > 0$ and $r_{jt}^0 < 0$, weighting each observation according to some kernel function $K(\frac{r_{jt}^0}{h})$. We recover the intercepts α_+ (using observations with positive values of r_{jt}^0) such that $\hat{y}_{jlt} = \hat{\alpha}_+ + \hat{\beta}_+ r_{jt}^0$, and α_- (using observations with negative values of r_{jt}^0) such that $\hat{y}_{jlt} = \hat{\alpha}_- + \hat{\beta}_- r_{jt}^0$. 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 $r_{jt}^0 \in [-h, h]$):

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

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 .¹⁵ 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 $\mathbb{E}[y_{jlt}(s)|r_{jt}]$ and $\mathbb{E}[y_{jlt}(s + 0.5)|r_{jt}]$ 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. 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.

¹⁵The inclusion of threshold fixed effects improves precision by controlling for the fact that the outcome level is different for each threshold.

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 Appendix Figure A-7.

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 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-8 shows that observable characteristics of GP practices are continuous through the thresholds. For our tests, we implement the RD methodology described above at the GP level, with five different types of GP practice characteristics on the left-hand side: (i) the fraction of patients that report having to wait one week or more for an appointment (in survey data) (ii) the number of months the GP has been active, (iii) a survey based outcome of patient trust, (iv) the QOF clinical score, and (v) the payments each GP receives per-patient from the NHS. In each panel, we report the t-test for the null that $\tau = 0$. We also show binned scatter-plots representing the means of each variable above and below the threshold, as well as estimates from local linear regressions. The average values of all are smooth through the threshold, with small t-statistics. 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-9, 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.

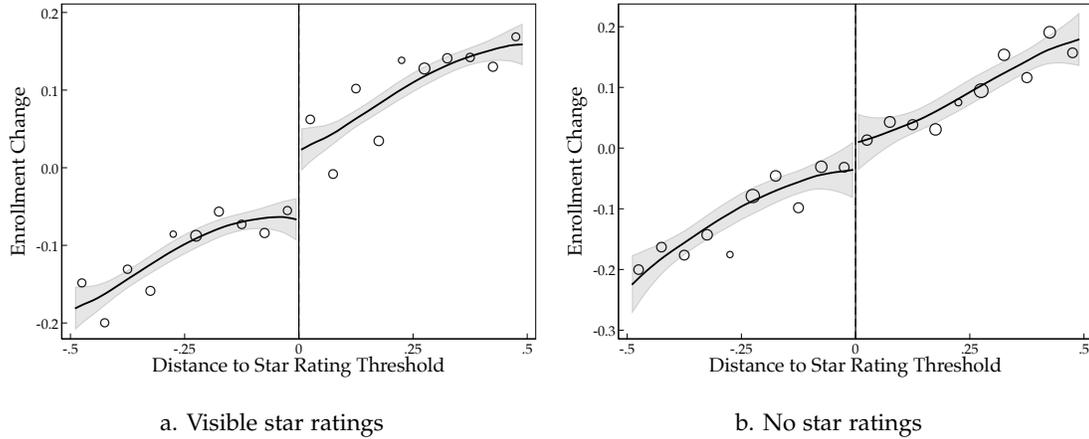
4.3 Regression Discontinuity Results

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. 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



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 kernel. 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 visible discontinuity when using data after January 2020.¹⁶ 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.

¹⁶We also do not find evidence of a discontinuity when considering only the first 3 months of 2020, to avoid the influence of the COVID-19 pandemic.

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 (0.058)	0.073 (0.034)	0.030 (0.105)	0.031 (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 (0.068)	0.058 (0.072)	-0.098 (0.140)	0.153 (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 kernel. 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.

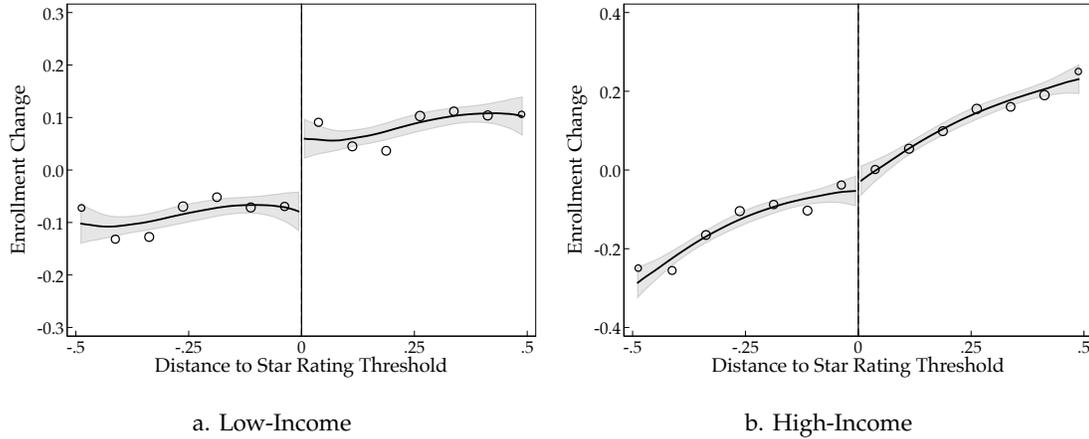
Differences by Income

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 of high versus low-income patient LSOAs.¹⁷

Figure 5 presents graphical evidence of the key result: the impacts of the star rating website are largely driven by patients in 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 LSOAs with above median income. The regression results in panel b of

¹⁷Our 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



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 kernel. Shaded area shows 95% confidence interval.

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 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 provides insight into individual 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. In contrast, we see relatively flat slopes for low-income LSOAs. Overall, the high-income sample is more than twice as responsive to quality despite the fact that they do not respond at the threshold. This can be seen by noting that the change in demand between a half star below the threshold and half star above the threshold is double the magnitude for the high-income sample. In the context of our model, these results suggests that both low and high-income individuals value quality, but that high-income individuals have more precise information from sources other than the rating website.

In Appendix Figure A-10 we show evidence that similar patterns hold across other mea-

asures 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-11, which shows separate plots by income quartile. The lowest income quartiles responds most at the star rating thresholds.

Robustness of 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.

Individuals do not frequently switch their GP. We show that the patterns we observe are consistent when limiting our sample only to movers: individuals who relocated their primary residence to a different LSOA in the same month that they enrolled in a new GP practice. The results, displayed in Figure A-12 and Panel B of Table A-3, show a similar differential response by income, implying that differential inertia with respect to income is not driving our findings. Even for individuals that are forced to make an active choice, low-income individuals respond more to the star ratings than high-income individuals.

Finally, there is concern about capacity constraints that (i) differentially bind around the constraint or (ii) differentially affect individuals by income. It is important to note that it is illegal for GPs to discriminate when accepting new patients and compensation of GPs is such that there is likely little financial preference for certain types of patients. According to the NHS, less than 2% of GPs are officially closed to new patients due to having too many patients.¹⁸ There is still concern that some GPs informally do not accept new patients or have long wait times, dissuading new patients from registering. To address concerns that capacity constraints or wait times are driving the differential effect of the star ratings, we limit the sample to high and low income patients living in close proximity. These individuals face

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

a similar choice set with similar potential capacity constraints or wait times. In addition, we have a specification in which we remove GPs with consecutive months with the same total enrollment, indicating potential capacity constraints. The RD results by income using these alternative samples, presented in Appendix Table A-4, are very similar to the baseline specification.

4.4 Alternative Identification Strategy: Panel Regressions

As a final piece of reduced form evidence, we show that enrollment responds to *within*-GP variation in rounded star ratings. Because individual reviews are added and subtracted each quarter, a GP practice’s star rating may differ over time even if underlying quality remains unchanged. As a result, we are able to estimate two-way fixed effects specifications that account for all time-invariant practice level factors. Specifically, we estimate:

$$y_{jlt} = \gamma s_{jt} + X_{jt}'\beta + \alpha_j + \delta_t + \epsilon_{jlt} \quad (6)$$

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.¹⁹ We also control for time-varying characteristics of the GP, X_{jt} , including age of the GP practice, age squared, and the number of practitioners in the GP practice. The main coefficient of interest is γ , the impact of an increase in rounded star ratings. As in our baseline RD specification, the outcome of interest is the quarterly change in enrollment at the GP-LSOA level and standard errors are clustered at the GP practice level.

The results, presented in Appendix Table A-5, are consistent with our RD strategy. In the full sample, star ratings have a positive and significant effect on quarterly change in enrollment. In Column 2 of Appendix Table A-5 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

¹⁹For 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.

responsive.

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. 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. However, 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. In order to separately identify the role of preferences and information in driving demand, previous literature has leveraged variation in information (Brown 2019; Allende et al. 2019; Brown and Jeon 2022). We explicitly incorporate the RD variation in information due to rounded star ratings in the empirical model. In addition to allowing for heterogeneous preferences and information, the model accounts for the presence of consumer inertia in health provider choice, which has been shown to be important in a variety of settings (e.g., Raval and Rosenbaum 2018; Shepard 2022).²⁰

5.1 Empirical Model Setup

If an individual makes an active choice, we assume utility follows the specification outlined in Section 3.1 with a few modifications. Posterior beliefs about the quality of GP j for

²⁰A 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).

individual i in LSOA ℓ and quarter t is

$$\mathbb{E}[r_{jt}|\tilde{r}_{ijt}, s_{jt}] = \alpha_{\ell t}\tilde{r}_{ijt} + (1 - \alpha_{\ell t})\mathbb{E}[r_{jt}|s_{jt}].$$

where the prior signal is $\tilde{r}_{ijt} = r_{jt} + \epsilon_{ijt}$. Therefore, utility from choosing GP $j \in \mathcal{J}_{\ell t}$ is given by

$$\mathbb{E}[u_{i\ell jt}] = \beta_{1\ell}\mathbb{E}[r_{jt}|\tilde{r}_{ijt}, s_{jt}] + f(d_{\ell j}, X_{\ell t}^d; \beta_2) + X_{jt}'\beta_3 + \zeta_j + v_{i\ell jt}.$$

For the estimation, we use the two-year rolling average of reviews for r_{jt} and apply a shrinkage procedure to $\mathbb{E}[r_{jt}|s_{jt}]$ that accounts for the variance of individual review scores and the number of reviews for each GP over the past two years. In addition to utility being a function of perceived quality as in Section 3.1, equation 5.1 includes preferences over other GP characteristics. In particular, utility depends on $f(d_{\ell j}, X_{\ell t}^d; \beta_2)$, a function of distance $d_{\ell j}$ from LSOA ℓ and j interacted with LSOA-level income, age and health.²¹ 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. ζ_j are GP fixed effects capturing amenities that are observed by the individual but not the researcher (e.g., convenient parking). We let preference for quality vary with (the z-score of) LSOA level income $\beta_{1\ell} = \beta_{10} + \beta_{11}I_{\ell}$.

The composite error, $v_{i\ell jt}$, incorporates both individual i 's (normalized) taste shock and the error, ϵ_{ij} , from the individuals private signal about the quality of GP j . Following the theoretical model, the variance of the composite error is given by $\text{Var}[v_{i\ell jt}] = \sigma_{v_{\ell}}^2 = \beta_{1\ell}^2 \alpha_{\ell t}^2 \sigma_{ij}^2 + 1$, where σ_{ij}^2 determines the precision of the individual's private signal. 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:

$$\sigma_{ij} = \exp[\gamma_0 + \gamma_1 I_{\ell}]. \quad (7)$$

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

²¹In the estimation, we use the functional form $f(d_{\ell j}, X_{\ell t}^d; \beta_2) = \exp((d_{\ell j} X_{\ell t}^d)' \beta_2)$, where $X_{\ell t}^d$ is a vector that includes a constant, LSOA income (z-score), LSOA average age, and LSOA average self-reported health status.

$$\alpha_{\ell t} = \frac{\sigma_s^2}{\exp[\gamma_0 + \gamma_1 I_\ell]^2 + \sigma_s^2}. \quad (8)$$

The variance of the information contained in star ratings, σ_s^2 , 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 $\varphi_{\ell t}$ that depends on observable characteristics of the LSOA. 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

$$\varphi_{\ell t} = \frac{\exp[X_{\ell t}^a \theta]}{1 + \exp[X_{\ell t}^a \theta]}$$

Given these specifications, we may write the market share for provider j within LSOA l in quarter t as²²:

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

Individuals who re-optimize have relatively standard logit choice probabilities. However, one complication is that the variance of the composite error differs across individuals due to heterogeneity in the precision of their private signal. The term $1/\tilde{\sigma}_{v_\ell}(\beta_{1\ell}, \gamma)$ normalizes the scale of the error. Individuals who do not re-optimize stay in the same practice as they were in $t - 1$.

5.2 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.

²²We let $\tilde{\sigma}_{v_\ell}^2 = \frac{6}{\pi^2} \sigma_{v_\ell}^2$.

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 $S_{\ell jt-1}$, which are data), we recover the market shares $\tilde{S}_{\ell jt}(\Phi)$ implied by the model. 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 of 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 well as the slopes on either side of the threshold. We run this separately for high and low-income LSOAs, and summarize the coefficients as a vector $\tilde{\tau}_I(\Phi)$, for $I \in \{\text{high}, \text{low}\}$.²³ When computing the RD moments, we use the same sample restrictions used for the results in Section 4.3, but we use the full sample when computing the remaining moments.

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:

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

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 k th element of this vector), the moments are

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

²³In practice, estimates from the parametric and non-parametric RD model are similar. See Appendix Table A-6.

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:

$$\hat{\Phi} = \operatorname{argmin}_{\Phi} M'WM.$$

where W is a positive definite weighting matrix. We compute standard errors by numerically computing $\frac{\partial M}{\partial \Phi_i}$ 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 (ζ_j), the preference for quality as measured by the reviews, and the precision of individuals' priors. Without targeting the RD estimates, ζ_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.

5.3 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-6). As in the baseline results, low-income individuals respond more at the star rating thresholds than high-income individuals. In ad-

dition, 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, $\mathcal{J}_{\ell t}$, 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.

The parameters governing inertia and preferences for other observables are reasonable. For example, patients dislike GP practices that are far from their 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.

5.4 Model Fit and Counterfactuals

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.

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 3, 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

Table 3
Estimates for GP Demand Model

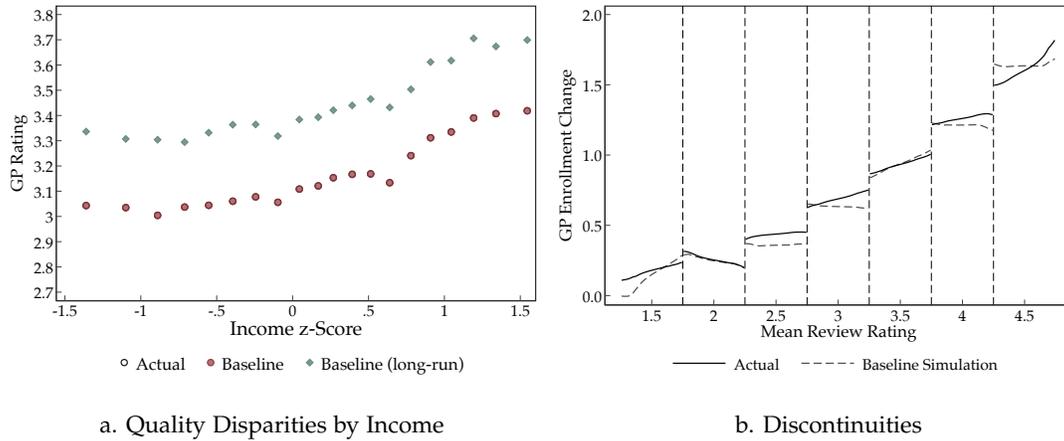
	Estimate	SE
<i>Inertia</i>		
Constant	-4.7864	(0.2295)
Income	0.0373	(0.0058)
Age	-0.0179	(0.0012)
Health	0.0977	(0.0472)
<i>Private Signal Variance ($\sigma_{\epsilon_t}^2$)</i>		
Constant	-1.2526	(0.0694)
Income	-0.9932	(0.0772)
<i>GP Quality</i>		
Constant	0.9002	(0.0071)
Income	0.0382	(0.0044)
<i>Distance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

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

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.

Figure 6
Model Fit: Matching Disparities and Regression Discontinuities



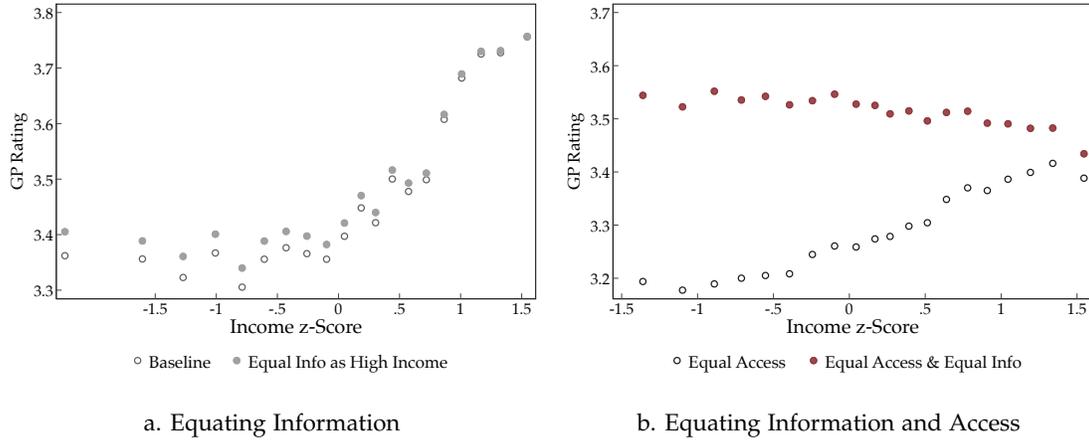
Notes: Panel a shows average GP quality by income computed from (i) the data (ii) our model, and (iii) the long run implied by our model (allowing all individuals to switch). Panel b plots enrollment changes within rounded star rating bins both based on the data and implied by our model.

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 the fraction of the observed income-quality gradient that can be explained by information itself.

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 providing this information to all individuals 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).

Figure 7
Counterfactual Experiments



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.

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

Table 4
Summary of Counterfactuals

	Baseline		Equal Info		Percent Change in Correlation
	Low-Income	High-Income	Low-Income	High-Income	
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: Baseline columns refer to long-run simulations (without inertia) 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.

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 eliminates 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.

A caveat to these counterfactuals is that they assume a fixed supply-side. While this assumption is reasonable in the short-run, it is possible that the endogenous entry, exit, or other responses of GP practices could alter our conclusions in the long-run. However, the consequences of a supply-side response for the gradient itself are ex-ante unclear. While a model of the supply-side 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

We provide regression discontinuity based evidence that information is an important driver of disparities in health care quality. Our analysis focuses on GP practice choice, where access to quality may have important implications for health care costs and health outcomes. While high-income patients choose better GPs on average, we find that low-income patients are significantly more responsive to online star ratings. High income individuals may be able to more easily research GP quality or have greater access to recommendations from their social networks. Conversely, low income individuals rely on information from star ratings.

We estimate a structural model in which individuals have heterogeneous private infor-

mation 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. Our counterfactuals show a significant reduction in the income-quality gradient, indicating that differential information plays an important role in driving inequality in health care quality along with differential access. Differences in preferences play a negligible role.

In many settings, there is been debate about allowing individuals the freedom to choose providers. While choice-based reforms can increase demand for high quality hospitals and GPs on average, with information gaps, such reforms may exacerbate inequality. Our findings suggest that both information and access are necessary to overcome existing disparities in healthcare quality.

References

- AHRQ, "National Healthcare Quality and Disparities Report," 2021.
- Allende, Claudia, "Competition under social interactions and the design of education policies," *Working Paper*, 2022.
- , Francisco Gallego, Christopher Neilson et al., "Approximating the equilibrium effects of informed school choice," Technical Report 2019.
- Anderson, Michael and Jeremy Magruder, "Learning from the crowd: Regression discontinuity estimates of the effects of an online review database," *The Economic Journal*, 2012, 122 (563), 957–989.
- Artmann, Elisabeth, Hessel Oosterbeek, and Bas Klaauw, "Do doctors improve the health care of their parents? Evidence from admission lotteries," 2019.
- Balarajan, Yarlani, 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.
- Bronnenberg, Bart J, Jean-Pierre Dubé, Matthew Gentzkow, and Jesse M Shapiro, "Do pharmacists buy Bayer? Informed shoppers and the brand premium," *The Quarterly Journal of Economics*, 2015, 130 (4), 1669–1726.
- Brown, Zach Y, "An empirical model of price transparency and markups in health care," *Working Paper*, 2019.
- and Jihye Jeon, "Endogenous information and simplifying insurance choice," *Working Paper*, 2022.
- Calonico, Sebastian, Matias D Cattaneo, and Rocio Titiunik, "Robust data-driven inference in the regression-discontinuity design," *The Stata Journal*, 2014, 14 (4), 909–946.
- , —, Max H Farrell, and Rocio Titiunik, "Regression discontinuity designs using covariates," *Review of Economics and Statistics*, 2019, 101 (3), 442–451.
- Cattaneo, Matias D, Michael Jansson, and Xinwei Ma, "Manipulation testing based on density discontinuity," *The Stata Journal*, 2018, 18 (1), 234–261.
- , Nicolás Idrobo, and Rocío Titiunik, *A practical introduction to regression discontinuity designs: Foundations*, Cambridge University Press, 2019.
- Chartock, Benjamin L, "Quality Disclosure, Demand, and Congestion: Evidence from Physician Ratings," 2021.
- Chen, Yiqun, Petra Persson, and Maria Polyakova, "The roots of health inequality and the value of intra-family expertise," Technical Report, National Bureau of Economic Research 2019.
- Chevalier, Judith A and Dina Mayzlin, "The effect of word of mouth on sales: Online book reviews," *Journal of Marketing Research*, 2006, 43 (3), 345–354.
- 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 Micheltore, 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.
- Greaves, Felix, Utz J Pape, Dominic King, Ara Darzi, Azeem Majeed, Robert M Wachter, and Christopher Millett**, "Associations between internet-based patient ratings and conventional surveys of patient experience in the English NHS: an observational study," *BMJ Quality & Safety*, 2012, 21 (7), 600–605.
- , **Utz J. Pape, Dominic King, Ara Darzi, Azeem Majeed, Robert M. Wachter, and Christopher Millett**, "Associations Between Web-Based Patient Ratings and Objective Measures of Hospital Quality," *Archives of Internal Medicine*, 03 2012, 172 (5), 435–436.
- 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.

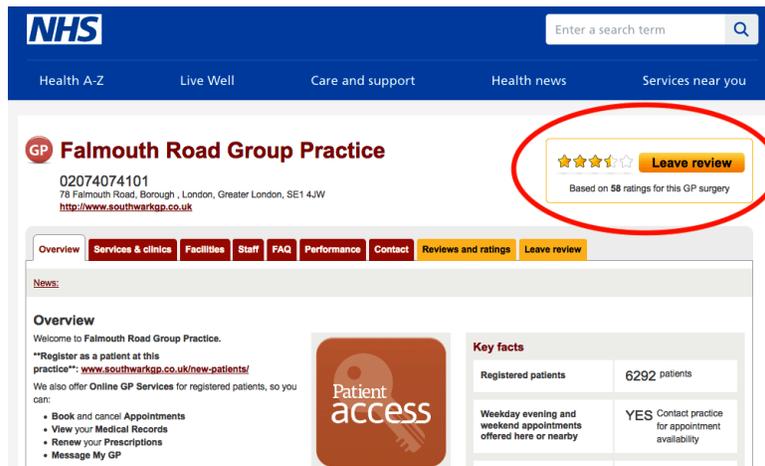
- Ho, Kate, Joseph Hogan, and Fiona Scott Morton**, "The impact of consumer inattention on insurer pricing in the Medicare Part D program," *The RAND Journal of Economics*, 2017, 48 (4), 877–905.
- Hortaçsu, Ali, Seyed Ali Madanizadeh, and Steven L Puller**, "Power to choose? An analysis of consumer 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.
- Kummer, Michael E, Ulrich Laitenberger, Cyrus E Rich, Danny R Hughes, and Turgay Ayer**, "Healthy reviews! Online physician ratings reduce healthcare interruptions," Technical Report, ZEW Discussion Papers 2021.
- Lewis, Gregory and Georgios Zervas**, "The welfare impact of consumer reviews: A case study of the hotel industry," Technical Report, Working Paper 2016.
- Luca, Michael**, "Reviews, reputation, and revenue: The case of Yelp. com," 2016. Harvard Business School Working Paper.
- **and Georgios Zervas**, "Fake it till you make it: Reputation, competition, and Yelp review fraud," *Management Science*, 2016, 62 (12), 3412–3427.
- 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," *British Medical Journal*, 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.
- Santos, Rita, Hugh Gravelle, and Carol Propper**, "Does quality affect patients' choice of doctor? Evidence from England," *The Economic Journal*, 2017, 127 (600), 445–494.
- Scobie, Sarah and Jessica Morris**, "Quality and inequality: digging deeper," Technical Report, Nuffield Trust January 2020.
- Shepard, Mark**, "Hospital network competition and adverse selection: evidence from the Massachusetts health insurance exchange," *American Economic Review*, 2022, 112 (2), 578–615.
- van Doorslaer, Eddy, Cristina Masseria, Xander Koolman et al.**, "Inequalities in access to medical care by income in developed countries," *Cmaj*, 2006, 174 (2), 177–183.
- , **Xander Koolman, and Andrew M Jones**, "Explaining income-related inequalities in doctor utilisation in Europe," *Health economics*, 2004, 13 (7), 629–647.
- Werner, Rachel M, Edward C Norton, R Tamara Konetzka, and Daniel Polsky**, "Do consumers respond to publicly reported quality information? Evidence from nursing homes," *Journal of Health Economics*, 2012, 31 (1), 50–61.

Online Appendix

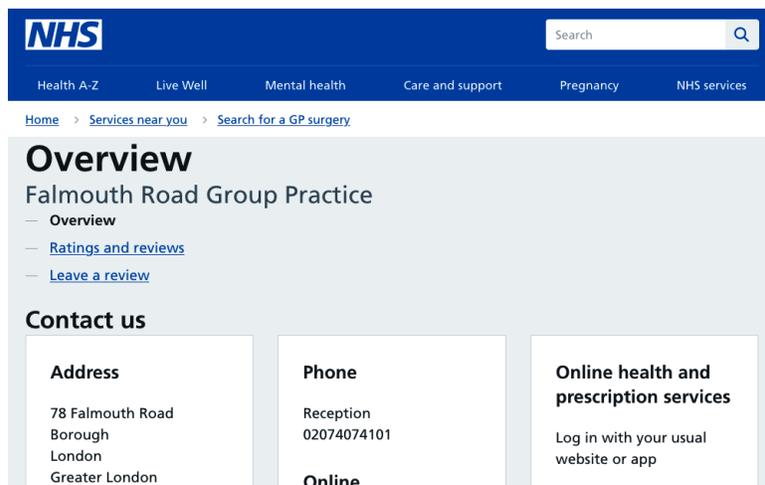
A Appendix Figures

Figure A-1
Example of the NHS Website with Visible Star Rating



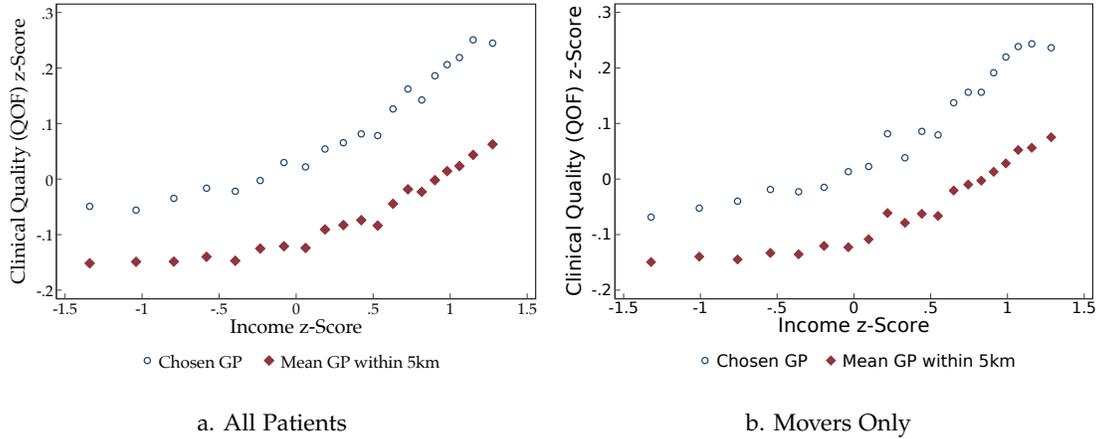
Notes: Shows NHS website for a GP practice prior to January 2020 (with visible star ratings).

Figure A-2
Example of the NHS Website without Star Rating



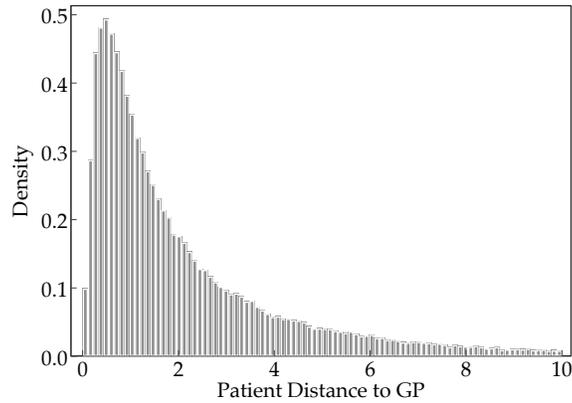
Notes: Shows NHS website for a GP practice after January 2020 when star ratings were removed from the website.

Figure A-3
Relationship between Income and Objective Measure of GP Quality



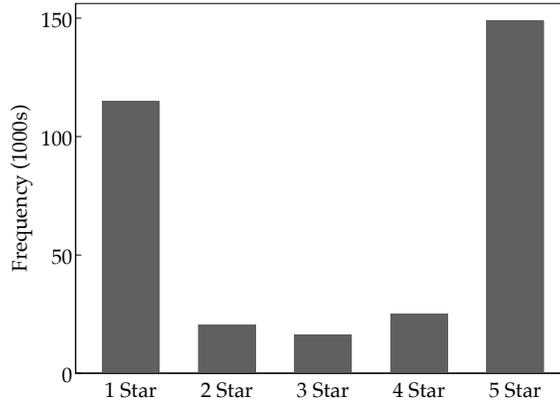
Notes: Binscatter plots show relationship between the z-score of patient LSOA income and the quality of the chosen GP and the mean quality of GPs within 5km. Quality is measured using the Quality Outcome Framework (QOF) clinical quality score, an objective measure of GP quality used for performance pay.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

Figure A-4
Histogram of Distance to GP



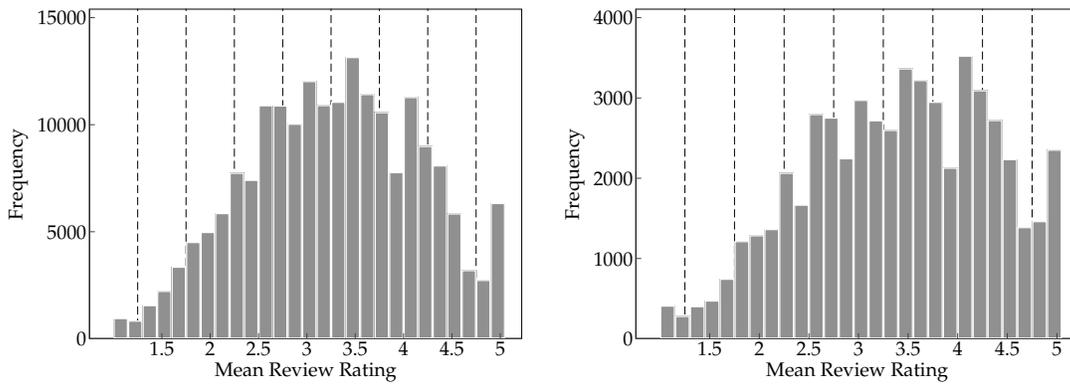
Notes: Chart shows histogram of individual distance to GPs.

Figure A-5
Histogram of Individual Reviews



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

Figure A-6
Histogram of Average GP Reviews

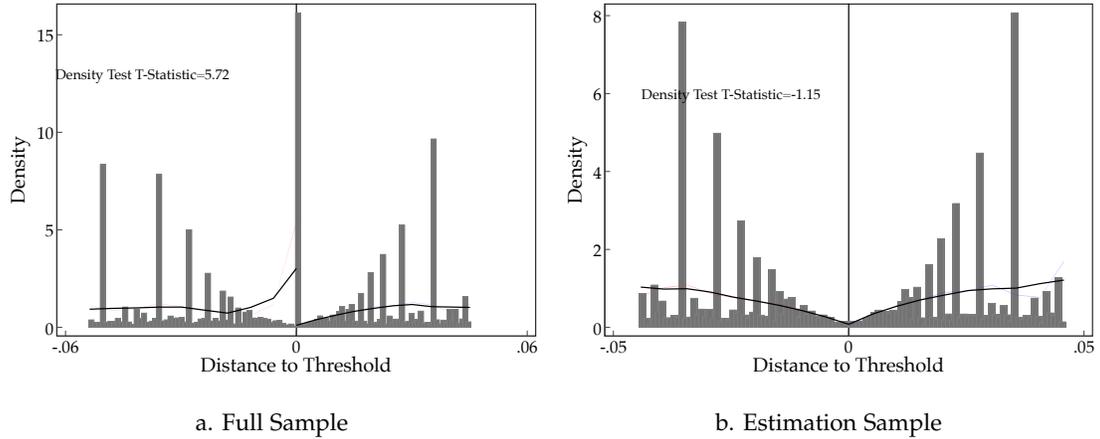


a. Visible star ratings

b. No star ratings

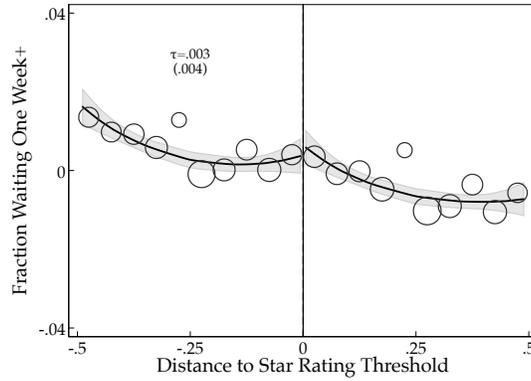
Notes: Chart shows histogram of average reviews for the period when the NHS Choices website displayed star ratings (“visible star ratings”) and the period when the website did not display star ratings (“no star ratings”). The NHS calculated average reviews using the running average of individual reviews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

Figure A-7
Density Tests of GP Average Reviews Around Star Rating
Thresholds

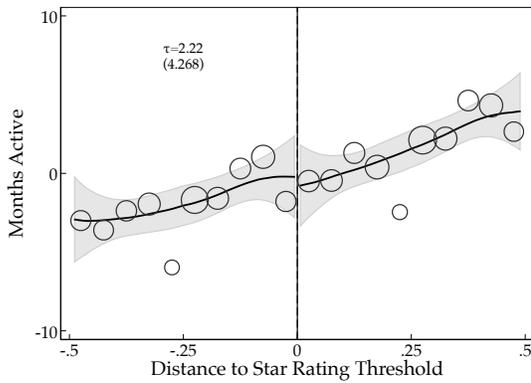


Notes: Figures show the histograms and polynomial density estimates for practices above and below star rounding thresholds. The unrestricted robust bias-corrected t-statistics following Cattaneo et al. (2018) are shown in the upper left hand corner. For the estimation sample, there is no evidence of bunching across the threshold.

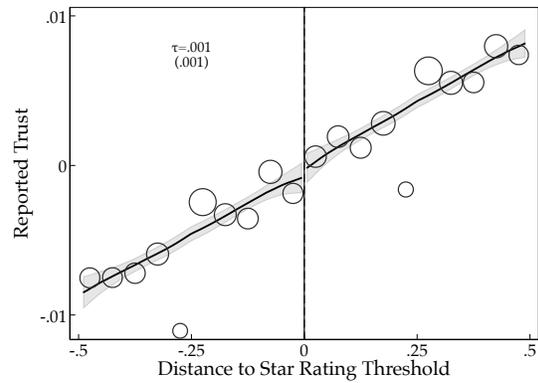
Figure A-8
Smoothness of GP-level Covariates Around Threshold



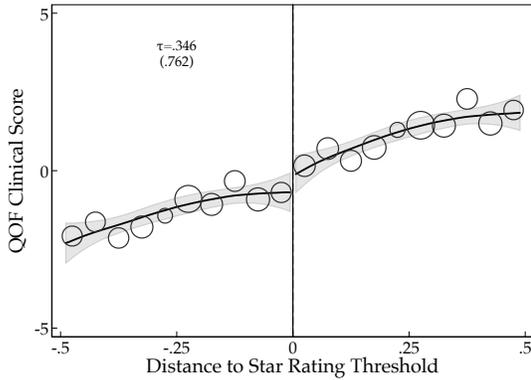
a. Waiting Times



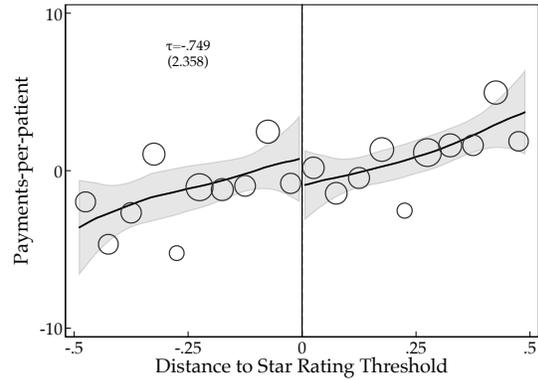
b. Months Active



c. Confidence and Trust (Survey)



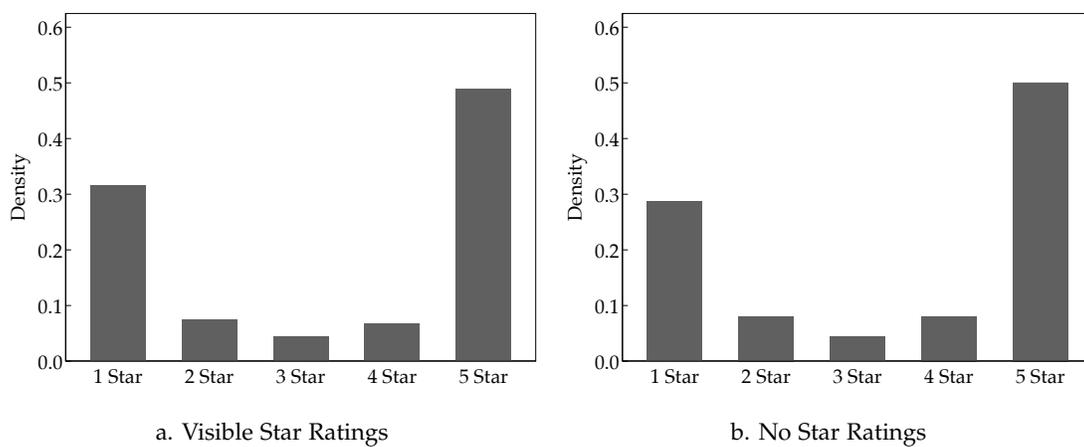
d. QOF Clinical Score



e. Payments Per-Patient

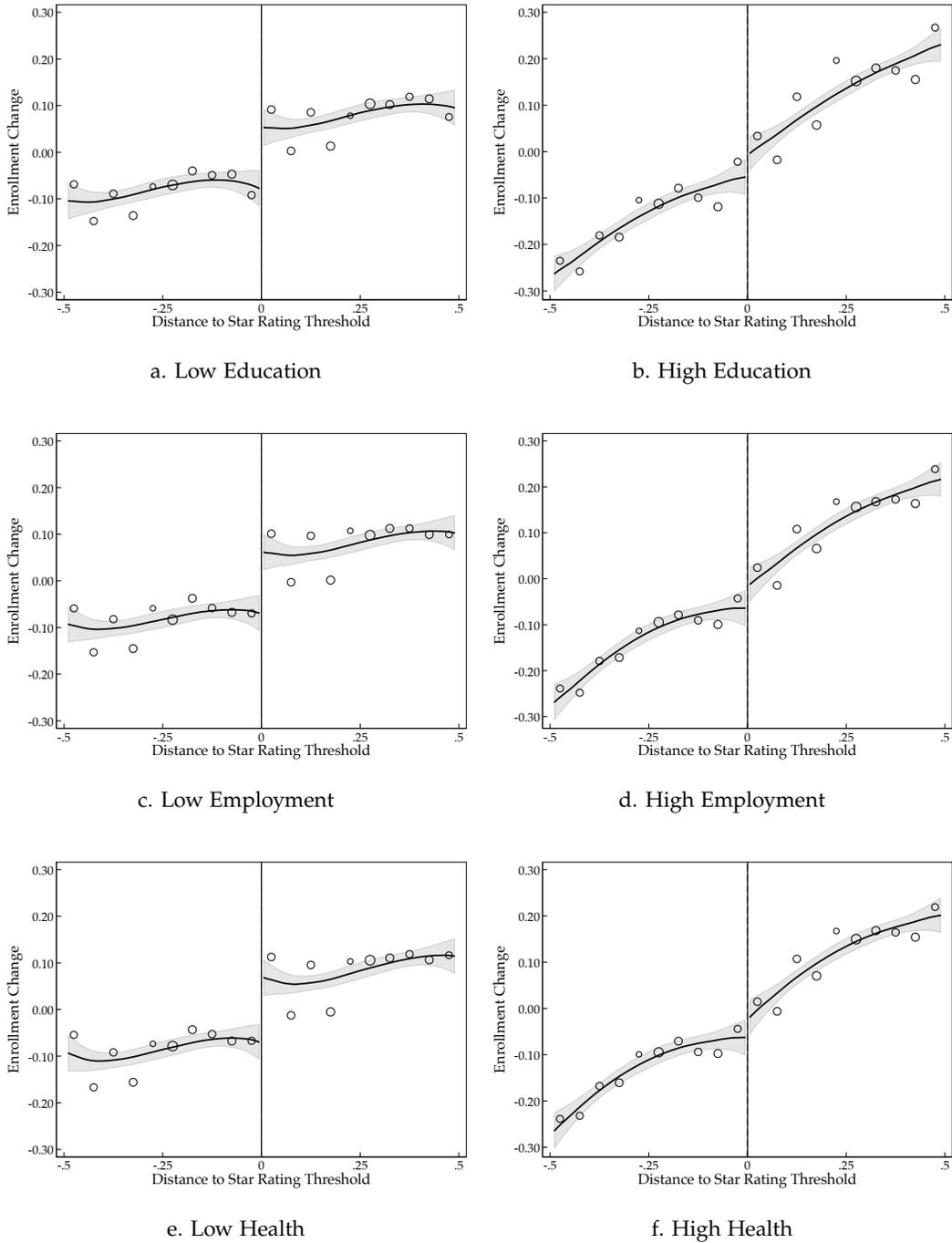
Notes: Changes at the threshold in GP-level observables. Circle size corresponds to the number of observations in each bin. Fitted line is from a local linear regression with a triangular kernel. Shaded area shows 95% confidence intervals.

Figure A-9
Distribution of Individual Reviews Around Website Change



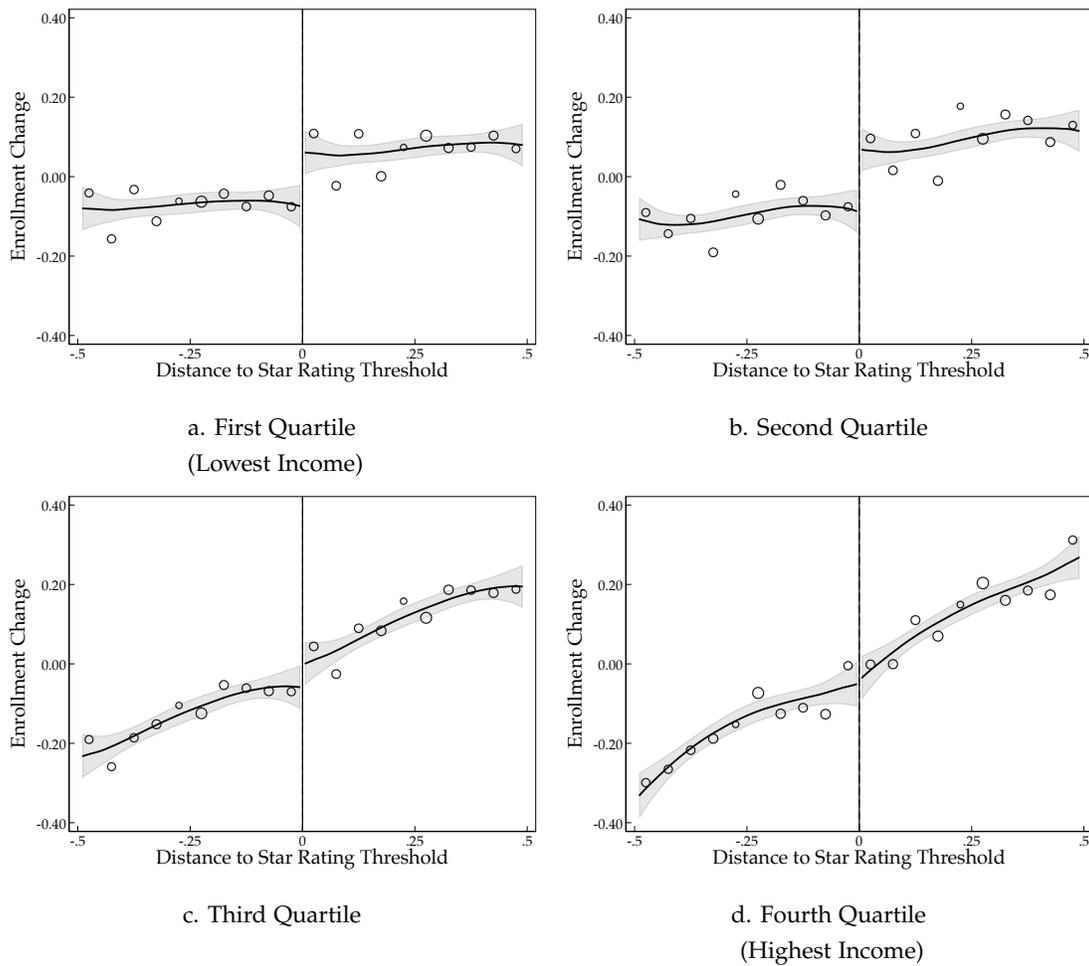
Notes: Sample is individual reviews a month before and after the website change in January 2020. We do not find a statistically significant difference in the distributions. The Chi-squared value is 5.6556 (p-value 0.226).

Figure A-10
 Effect of Star Rating Threshold on Enrollment Change
 Additional Heterogeneity Analysis



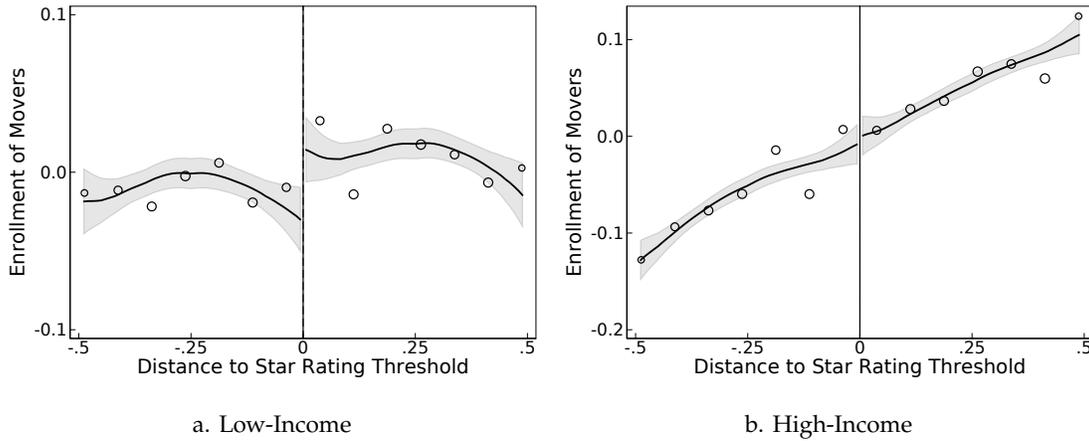
Notes: Chart shows mean enrollment change around threshold for star ratings by quartile of LSOA income of patients for the period in which star ratings were visible. The size of the circles corresponding to the number of observations in each bin. The fitted line is from a local linear regression using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

Figure A-11
 Effect of Star Rating Threshold on GP Enrollment
 by Income Quartile



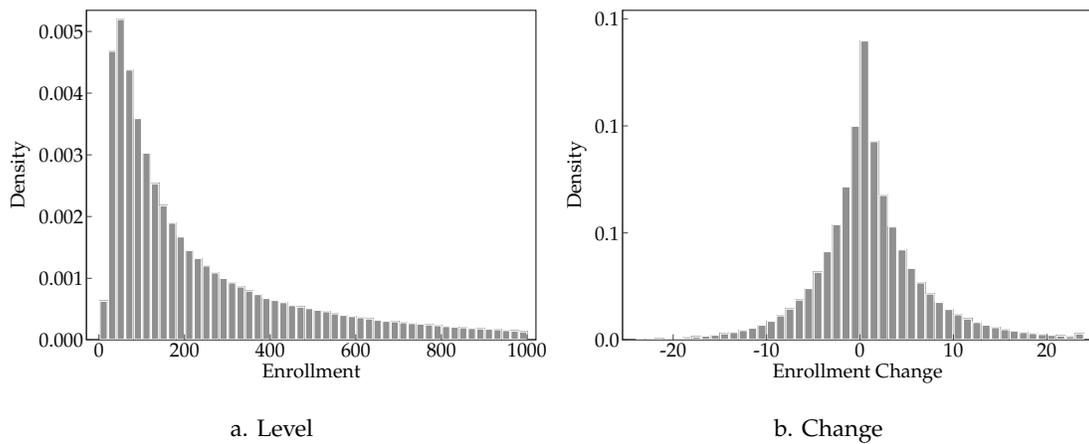
Notes: Chart shows mean enrollment change around threshold for star ratings by quartile of LSOA income of patients for the period in which star ratings were visible. The size of the circles corresponding to the number of observations in each bin. The fitted line is from a local linear regression using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

Figure A-12
 Effect of Star Rating Threshold on GP Enrollment by Income
 Sample of Movers



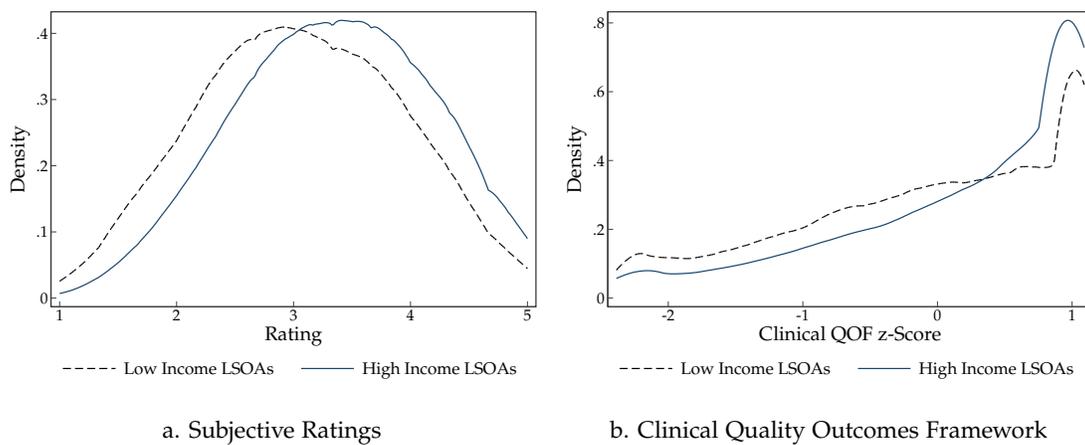
Notes: Chart shows mean enrollment change for movers around threshold for star ratings, splitting at median LSOA income, for the period in which star ratings were visible. Movers are defined as individuals who relocated to a new LSOA and registered with a new GP practice in the same month. The size of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

Figure A-13
 Histogram of GP Enrollment by LSOA-Quarter



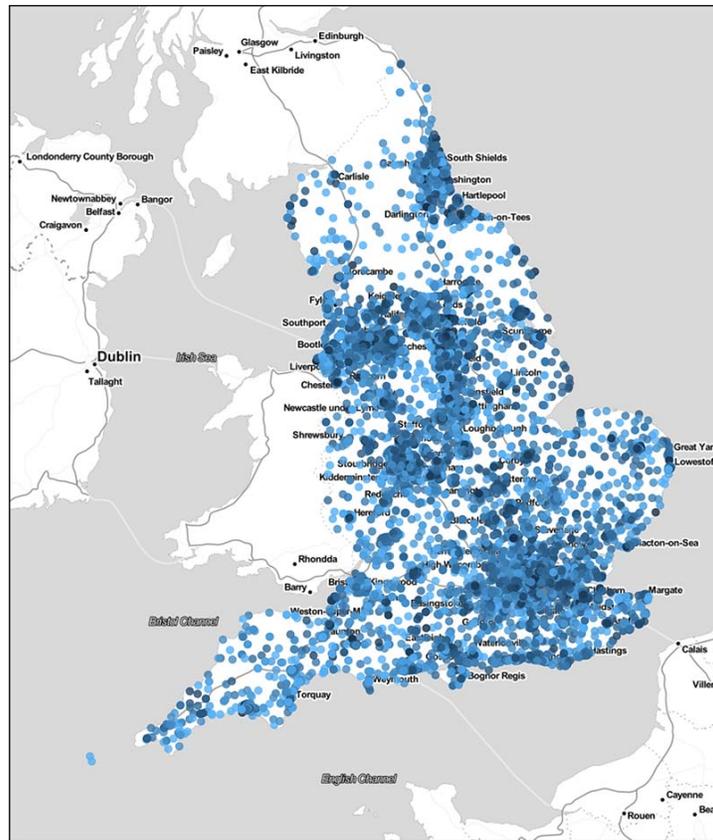
Notes: Chart shows histogram of enrollment by LSOA-quarter and quarterly change in enrollment. Enrollment above 1,000 not shown in panel (a). Absolute change in enrollment greater than 25 not shown in panel (b).

Figure A-14
Density of GP Quality in Choice Set by Income



Notes: Charts shows histogram of GP ratings and GP clinical Quality Outcome Framework (QOF) 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-15
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
<i>Patient Surveys:</i>						
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
<i>Quality and Outcomes Framework:</i>						
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
<i>Prescription Drugs:</i>						
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
RD Estimates for Full Sample
Robustness to Alternative Specifications and Bandwidths

Alternative Specifications					
	Epanechnikov Kernel	No Min. # Reviews	Include $r_{jt} = c_s$	No Covariates	% Change Enroll.
Estimate	0.135** (0.060)	0.112** (0.052)	0.097** (0.049)	0.082 (0.051)	0.276*** (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** (0.067)	0.110** (0.045)	0.086** (0.037)	0.072** (0.034)	0.067** (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
RD Estimates by Income
Robustness to Alternative Specifications and Bandwidths

Panel a: Alternative Specifications										
	Epanechnikov Kernel		No Min. # Reviews		Include $r_{jt} = c_s$		No Covariates		% Change Enroll.	
	Low Inc.	High Inc.	Low Inc.	High Inc.	Low Inc.	High Inc.	Low Inc.	High Inc.	Low Inc.	High Inc.
Estimate	0.213*** (0.073)	0.059 (0.074)	0.185*** (0.068)	0.057 (0.071)	0.166*** (0.058)	0.035 (0.062)	0.161** (0.067)	0.005 (0.060)	0.463*** (0.158)	0.063 (0.065)
Bandwidth	0.12	0.11	0.15	0.13	0.17	0.13	0.15	0.18	0.12	0.16
N	420,139	397,821	547,867	498,345	634,003	499,627	560,213	636,077	419,931	591,464

Panel b: Movers and Alternative Bandwidths										
	Movers Only		Bandwidth=0.1		Bandwidth=0.3		Bandwidth=0.4		Bandwidth=0.5	
	Low Inc.	High Inc.	Low Inc.	High Inc.	Low Inc.	High Inc.	Low Inc.	High Inc.	Low Inc.	High Inc.
Estimate	0.118** (0.059)	0.000 (0.059)	0.223*** (0.084)	0.095 (0.082)	0.145*** (0.047)	0.027 (0.045)	0.119*** (0.042)	0.027 (0.040)	0.111*** (0.038)	0.024 (0.037)
Bandwidth	0.10	0.11	0.10	0.10	0.30	0.30	0.40	0.40	0.50	0.50
N	336,678	391,313	347,252	351,372	1,076,997	1,091,008	1,424,869	1,452,231	1,742,940	1,774,703

Notes: Panel a shows robustness for RD specifications. The first set of columns employs the epanechnikov kernel, the second includes GP practices with fewer than 5 reviews, the third includes observations with the index r_{jt} exactly equal to a rounding threshold, the fourth excludes all covariates, and the fifth uses the percent change in enrollment at the LSOA-GP level as the dependent variable. Panel b shows robustness to limiting to movers and the bandwidth choice. Movers are defined as individuals who relocated to a new LSOA and registered with a new GP practice in the same month. Sample period is when stars were visible, and standard errors clustered at the GP level are included in parenthesis. Except where otherwise noted, the dependent variable is quarterly enrollment change for an LSOA-GP, 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
RD Estimates by Income
Robustness Addressing Potential Capacity Constraints

	Similar Choice Set		No GPs with Static Enrollment	
	Low Inc.	High Inc.	Low Inc.	High Inc.
Estimate	0.154 (0.073)	0.097 (0.081)	0.190 (0.073)	0.075 (0.076)
P-Value	0.036	0.229	0.009	0.321
CCT Robust P-Value	0.048	0.208	0.012	0.329
Bandwidth	0.14	0.13	0.15	0.13
N	383,745	308,467	489,335	410,061

Notes: Dependent variable is change in GP enrollment. The first two columns limit the sample to high (low) income LSOAs that are within 1km of a low (high) income LSOA, implying that low and high income individuals face a similar choice set. Columns 3 and 4 remove GPs with enrollment that does not change between at least two consecutive months within a year. Sample excludes quarterly GP observations with fewer than five reviews. Each regression controls for average review score using a local linear regression with triangular kernel. CCT optimal bandwidths are calculated for each specification. Low (high) income is defined as those in an LSOA with below (above) median income. Standard errors are in parentheses and clustered by GP.

Table A-5
Effect of Star Ratings on Enrollment Change
Panel Regression Estimates

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

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

Table A-6
Effect of Star Ratings on Enrollment Change
Simple Parametric Regression Discontinuity Estimates

	Visible Star Ratings	No Star Ratings	Visible Star Ratings		Visible Star Ratings London Only	
			Low Income	High Income	Low Income	High Income
Estimate	0.072 (0.028)	0.017 (0.049)	0.113 (0.035)	0.035 (0.034)	0.159 (0.064)	0.047 (0.063)
Distance from threshold	-0.031 (0.056)	0.099 (0.097)	-0.129 (0.069)	0.065 (0.068)	-0.216 (0.129)	0.039 (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: Results from OLS regression in which 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.